Figure 1

CGAGAGACGACAGAAGGTTACGGCTGCGAGAAGACGACAGAAGGGTCCAGAAAAA
GGAAAGTGCTGGAGGGGAGTGGGGACAAAAGCAGCGACCAAGTGAATGTCACTTC
AGTGACTGAGGCCAGGCAAAACGCGCGGGAAGGATTTTGTGTAGCTTGGGACCCTT
TCATAGACACTGATGACACGTTTACGCAAAATAGAAATTTGAGGAGAAACGCCTGG
GCCTTCGGAAAGGAGTGATTGATTAGTACTTGCAAGTTTAGGTGACTTTAAGGAGAA
CTAACTAATGTATACTATTGAGGGAGGAGGAAGAGCATTACAGAGTTTCCAGCAGC
AGCAGGAAAGCTTTGGTTAATTTGGAAATGGATGATAGCATTAAAATAACAGAAGC
GCCTCCAGGTCTCTGAAGCTTCAGTCCCCAGCTGAAAGCCAGAAAAGACTAAGCC
CACTAAGCCTTTTGATCCCTTTGGAAGCAAAGAACTTTCCTTCCCTGGGGTGAAGAC
TCTCCTCAGAAGATTTCCTGTCTCTGCCTATGTTACAAGAGGAATCAAAACCAAGAC
AGAAGAGCTCAGGATGCAGGTGAGAGGCAGGGAAGTCAGCGGCTTGTTGATCTCCC
TCACTGCTGTCTGCCTGGTGGTCACCCCTGGGAGCAGGGCCTGTCCTCGCCGCTGTG
CCTGCTATGTGCCCACAGAGGTGCACTGTACATTTCGGTACCTGACCTCCATCCCAG
ATGGCATCCCGGCCAATGTGGAACGAATAAATTTAGGATATAACAGCCTTACTAGAT
TGACAGAAAACGACTTTGATGGCCTGAGCAAACTGGAGTTACTCATGCTGCACAGT
AATGGCATTCACAGAGTCAGTGACAAGACCTTCTCGGGCTTGCAGTCCTTGCAGGTC
TTAAAAATGAGCTATAACAAAGTCCAAATCATTCGGAAGGATACTTTCTACGGACTC
GGGAGCTTGGTCCGGTTGCACCTGGATCACAACAACATTGAATTCATCAACCCTGAG
GCCTTTTATGGACTTACCTCGCTCCGCTTGGTACATTTAGAAGGAAACCGGCTCACA
AAGCTCCATCCAGACACATTTGTCTCATTAAGCTATCTCCAGATATTTAAAACCTCTT
TCATTAAGTACCTGTTCTTGTCTGATAACTTCCTGACCTCCCTCCCAAAAGAAATGGT
CTCCTACATGCCAAACCTAGAAAGCCTGTATTTGCATGGAAACCCATGGACCTGTGA
CTGCCATTTAAAGTGGTTGTCTGAGTGGATGCAGGGAAACCCAGATATAATAAAAT
GCAAGAAAGACAGAAGCTCTTCCAGTCCTCAGCAATGTCCCCTTTGCATGAACCCCA
GGATCTCTAAAGGCAGACCCTTTGCTATGGTACCATCTGGAGCTTTCCTATGTACAA
AGCCAACCATTGATCCATCACTGAAGTCAAAGAGCCTGGTTACTCAGGAGGACAAT
GGATCTGCCTCCACCTCACCTCAAGATTTCATAGAACCCTTTGGCTCCTTGTCTTTGA
ACATGACANANNTNTCTGGAAATAAGGCCGACATGGTCTGTAGTATCCAAAAGCCA
TCAAGGACATCACCAACTGCATTCACTGAAGAAAATGACTACATCATGCTAAATGC
GTCATTTTCCACAAATCTTGTGTGCAGTGTAGATTATAATCACATCCAGCCAGTGTG
GCAACTTCTGGCTTTATACAGTGACTCTCCTCTGATACTAGAAAGGAAGCCCCAGCT
TACCGAGACTCCTTCACTGTCTTCTAGATATAAACAGGTGGCTCTTAGGCCTGAAGA
CATTTTTACCAGCATAGAGGCTGATGTCAGAGCAGACCCTTTTTGGTTCCAACAAGA
AAAAATTGTCTTGCAGCTGAACAGAACTGCCACCACACTTAGCACATTACAGATCCA
GTTTTCCACTGATGCTCAAATCGCTTTACCAAGGGCGGAGATGAGAGCGGAGAGAC
TCAAATGGACCATGATCCTGATGATGAACAATCCCAAACTGGAACGCACTGTCCTGG
TTGGCGGCACTATTGCCCTGAGCTGTCCAGGCAAAGGCGACCCTTCACCTCACTTGG
AATGGCTTCTAGCTGATGGGAGTAAAGTGAGAGCCCCTTACGTTAGCGAGGATGGG

CGAATCCTAATAGACAAAAATGGGAAGTTGGAACTGCAGATGGCTGACAGCTTTGA
TGCAGGTCTTTACCACTGCATAAGCACCAATGATGCAGATGCGGATGTTCTCACATA
CAGGATAACTGTGGTAGAGCCCTATGGAGAAAGCACACATGACAGTGGAGTCCAGC
ACACAGTGGTTACGGGTGAGACGCTCGACCTTCCATGCCTTTCCACGGGTGTTCCAG
ATGCTTCTATTAGCTGGATTCTTCCAGGGAACACTGTGTTCTCTCAGCCATCAAGAG
ACAGGCAAATTCTTAACAATGGGACCTTAAGAATATTACAGGTTACGCCAAAAGAT
CAAGGTCATTACCAATGTGTGGCTGCCAACCCATCAGGGGCCGACTTTTCCAGTTTT
AAAGTTTCAGTTCAAAAGAAAGGCCAAAGGATGGTTGAGCATGACAGGGAGGCAG
GTGGATCTGGACTTGGAGAACCCAACTCCAGTGTTTCCCTTAAGCAGCCAGCATCTT
TGAAACTCTCTGCATCAGCTTTGACAGGGTCAGAGGCTGGAAAACAAGTCTCCGGTG
TACATAGGAAGAACAAACATAGAGACTTAATACATCGGCGGCGTGGGGATTCCACG
CTCCGGCGATTCAGGGAGCATAGGAGGCAGCTCCCTCTCTCTGCTCGGAGAATTGAC
CCGCAACGCTGGGCAGCACTTCTAGAAAAAGCCAAAAAGAATTCTGTGCCAAAAAA
GCAAGAAAATACCACAGTAAAGCCAGTGCCACTGGCTGTTCCCCTCGTGGAACTCA
CTGACGAGGAAAAGGATGCCTCTGGCATGATTCCTCCAGATGAAGAATTCATGGTTC
TGAAAACTAAGGCTTCTGGTGTCCCAGGAAGGTCACCAACTGCTGACTCTGGACCAG
TAAATCATGGTTTTATGACGAGTATAGCTTCTGGCACAGAAGTCTCAACTGTGAATC
CACAAACACTACAATCTGAGCACCTTCCTGATTTCAAATTATTTAGTGTAACAAACG
GTACAGCTGTGACAAAGAGTATGAACCCATCCATAGCAAGCAAAATAGAAGATACA
ACCAACCAAAACCCAATCATTATCTTTCCATCAGTAGCTGAAATTCGAGATTCTGCT
CAGGCAGGAAGAGCATCTTCCCAAAGTGCACACCCTGTAACAGGGGGAAACATGGC
TACCTATGGCCATACCAACACATATAGTAGCTTTACCAGCAAAGCCAGTACAGTCTT
GCAGCCAATAAATCCAACAGAAAGTTATGGACCTCAGATACCTATTACAGGAGTCA
GCAGACCTAGCAGTAGTGACATCTCTTCTCACACTACTGCAGACCCTAGCTTCTCCA
GTCACCCTTCAGGTTCACACACCACTGCCTCGTCTTTATTTCACATTCCTAGAAACAA
CAATACAGGTAACTTCCCCTTGTCCAGGCACTTGGGAAGAGAGAGGACAATTTGGA
GCAGAGGGAGAGTTAAAAACCCACATAGAACCCCAGTTCTCCGACGGCATAGACAC
AGGACTGTGAGGCCAGCAATCAAGGGACCTGCTAACAAAAATGTGAGCCAAGTTCC
AGCCACAGAGTACCCTGGGATGTGCCACACATGTCCTTCCGCAGAGGGGCTCACAG
TGGCTACTGCAGCACTGTCAGTTCCAAGTTCATCCCACAGTGCCCTCCCCAAAACTA
ATAATGTTGGGGTCATAGCAGAAGAGTCTACCACTGTGGTCAAGAAACCACTGTTAC
TATTTAAGGACAAACAAATGTAGATATTGAGATAATAACAACCACTACAAAATAT
TCCGGAGGGGAAAGTAACCACGTGATTCCTACGGAAGCAAGCATGACTTCTGCTCC
AACATCTGTATCCCTGGGGAAATCTCCTGTAGACAATAGTGGTCACCTGAGCATGCC
TGGGACCATCCAAACTGGGAAAGATTCAGTGGAAACAACACCACTTCCCAGCCCCC
TCAGCACACCCTCAATACCAACAAGCACAAAATTCTCAAAGAGGAAAACTCCCTTG
CACCAGATCTTTGTAAATAACCAGAAGAAGGAGGGGATGTTAAAGAATCCATATCA
ATTCGGTTTACAAAAGAACCCAGCCGCAAAGCTTCCCAAAATAGCTCCTCTTTTACC
CACAGGTCAGAGTTCCCCCTCAGATTCTACAACTCTCTTGACAAGTCCGCCACCAGC
TCTGTCTACAACAATGGCTGCCACTCAGAACAAGGGCACTGAAGTAGTATCAGGTG

CCAGAAGTCTCTCAGCAGGGAAGAAGCAGCCCTTCACCAACTCCTCTCCAGTGCTTC
CTAGCACCATAAGCAAGAGATCTAATACATTAAACTTCTTGTCAACGGAAACCCCCA
CAGTGACAAGTCCTACTGCTACTGCATCTGTCATTATGTCTGAAACCCAACGAACAA
GATCCAAAGAAGCAAAAGACCAAATAAAGGGGCCTCGGAAGAACAGAAACAACGC
AAACACCACCCCCAGGCAGGTTTCTGGCTATAGTGCATACTCAGCTCTAACAACAGC
TGATACCCCCTTGGCTTTCAGTCATTCCCCACGACAAGATGATGGTGGAAATGTAAG
TGCAGTTGCTTATCACTCAACAACCTCTCTTCTGGCCATAACTGAACTGTTTGAGAA
GTACACCCAGACTTTGGGAAATACAACAGCTTTGGAAACAACGTTGTTGAGCAAAT
CACAGGAGAGTACCACAGTGAAAAGAGCCTCAGACACACCACCACCACTCCTCAGC
AGTGGGGCGCCCCCAGTGCCCACTCCTTCCCCACCTCCTTTTACTAAGGGTGTGGTT
ACAGACAGCAAAGTCACATCAGCTTTCCAGATGACGTCAAATAGAGTGGTCACCAT
ATATGAATCTTCAAGGCACAATACAGATCTGCAGCAACCCTCAGCAGAGGCTAGCC
CCAATCCTGAGATCATAACTGGAACCACTGACTCTCCCTCTAATCTGTTTCCATCCAC
TTCTGTGCCAGCACTAAGGGTAGATAAACCACAGAATTCTAAATGGAAGCCCTCTCC
CTGGCCAGAACACAAATATCAGCTCAAGTCATACTCCGAAACCATTGAGAAGGGCA
AAAGGCCAGCAGTAAGCATGTCCCCCCACCTCAGCCTTCCAGAGGCCAGCACTCAT
GCCTCACACTGGAATACACAGAAGCATGCAGAAÁGAGTGTTTTTGATAAGAAACC
TGGTCAAAACCCAACTTCCAAACATCTGCCTTACGTCTCTCTACCTAAGACTCTATTG
AAAAAGCCAAGAATAATTGGAGGAAAGGCTGCAAGCTTTACAGTTCCAGCTAATTC
AGACGTTTTTCTTCCTTGTGAGGCTGTTGGAGACCCACTGCCCATCATCCACTGGACC
AGAGTTTCATCAGGANTTGAAATATCCCAAGGGACACAGAAAAGCCGGTTCCACGT
GCTTCCCAATGGCACCTTGTCCATCCAGAGGGTCAGTATTCAGGACCGTGGACAGTA
CCTGTGCTCTGCATTTAATCCACTGGGCGTAGACCATTTTCATGTCTCTTTGTCTGTG
GTTTTTACCCGGCAAGGATTTTGGACAGACATGTCAAGGAGATCACAGTTCACTTT
GGAAGTACTGTGGAACTAAAGTGCAGAGTGGAGGGTATGCCGAGGCCTACGGTTTC
CTGGATACTTGCAAACCAAACGGTGGTCTCAGAAACGGCCAAGGGAAGCAGAAAGG
TCTGGGTAACACCTGATGGAACATTGATCATCTATAATCTGAGTCTTTATGATCGTG
GTTTTTACAAGTGTGTGGCCAGCAACCCATCTGGCCAGGATTCACTGTTGGTTAAGA
TACAAGTCATCACAGCTCCCCCTGTCATTATAGAGCAAAAGAGGCAAGCCATCGTTG
GGGTTTTAGGTGGAAGTTTGAAACTGCCCTGCACTGCAAAAGGAACTCCCCAGCCTA
GTGTTCACTGGGTCCTTTATGATGGGACTGAACTAAAACCATTGCAGTTGACTCATT
CCAGATTTTTCTTGTATCCAAATGGAACTCTGTATATAAGAAGCATCGCTCCTTCAGT
GAGGGGCACTTATGAGTGCATTGCCACCAGCTCCTCAGGCTCAGAGAGAAGGGTAG
TGATTCTTACTGTGGAAGAGGGAGAGACAATCCCCAGGATAGAAACTGCCTCTCAG
AAATGGACTGAGGTGAATTTGGGTGAGAAATTACTACTGAACTGCTCAGCTACTGG
GGATCCAAAGCCTAGAATAATCTGGAGGCTGCCATCCAAGGCTGTCATCGACCAGT
GGCACAGAATGGGCAGCCGAATCCACGTCTACCCAAATGGATCCTTGGTGGTTGGG
TCAGTGACGGAAAAAGACGCTGGTGACTACTTATGTGTGGCAAGAAACAAAATGGG
AGATGACCTAGTCCTGATGCATGTCCGCCTGAGATTGACACCTGCCAAAATTGAACA
GAAGCAGTATTTTAAGAAGCAAGTGCTCCATGGGAAAGATTTCCAAGTTGACTGCA

AGGCCTCTGGCTCCCCTGTGCCTGAGGTATCCTGGAGTTTGCCTGATGGGACAGTGC
TCAACAATGTAGCCCAAGCTGATGACAGTGGCTATAGGACCAAGAGGTACACCCTT
TTCCACAATGGAACCTTGTATTTCAACAACGTTGGGATGGCAGAGGAAGGAGATTAT
ATCTGCTCTGCCCAGAACACCTTAGGGAAAGATGAGATGAAAGTCCACCTAACAGT
TCTAACAGCCATCCCACGGATAAGGCAAAGCTACAAGACCACCATGAGGCTCAGGG
CTGGAGAAACAGCTGTCCTTGACTGCGAGGTCACTGGGGAACCGAAGCCCAATGTA
TTTTGGTTGCTGCCTTCCAACAATGTCATTTCATTCTCCAATGACAGGTTCACATTTC
ATGCCAATAGAACTTTGTCCATCCATAAAGTGAAACCACTTGACTCTGGGGACTATG
TGTGCGTAGCTCAGAATCCTAGTGGGGATGACACTAAGACATACAAACTGGACATT
GTCTCTAAACCTCCATTAATCAATGGCCTGTATGCAAACAAGACTGTTATTAAAGCC
ACAGCCATTCGGCACTCCAAAAAATACTTTGACTGCAGAGCAGATGGGATCCCATCT
TCCCAGGTCACGTGGATTATGCCAGGCAATATTTTCCTCCCAGCTCCATACTTTGGA
AGCAGAGTCACGGTCCATCCAAATGGAACCTTGGAGATGAGGAACATCCGGCTTTC
TGACTCTGCGGACTTCACCTGTGTGGTTCGGAGCGAGGGAGGAGAGAGTGTGTTGGT
AGTGCAGTTAGAAGTCCTAGAAATGCTGAGAAGACCAACATTCAGAAACCCATTCA
ACGAAAAAGTCATCGCCCAAGCTGGCAAGCCCGTAGCACTGAACTGCTCTGTGGAT
GGGAACCCCCCACCTGAAATTACCTGGATCTTACCTGACGGCACACAGTTTGCTAAC
AGACCACACAATTCCCCGTATCTGATGGCAGGCAATGGCTCTCTCATCCTTTACAAA
GCAACTCGGAACAAGTCAGGGAAGTATCGCTGTGCAGCCAGGAATAAGGTTGGCTA
CATCGAGAAACTCATCCTGTTAGAGATTGGGCAGAAGCCAGTCATTCTGACATACGA
ACCAGGGATGGTGAAGAGCGTCAGTGGGGAACCGTTATCACTGCATTGTGTGTCTG
ATGGGATCCCCAAGCCAAATGTCAAGTGGACTACACCGGGTGGCCATGTAATCGAC
AGGCCTCAAGTGGATGGAAAATACATACTGCATGAAAATGGCACGCTGGTCATCAA
AGCAACAACAGCTCACGACCAAGGAAATTATATCTGTAGGGCTCAAAACAGTGTTG
GCCAGGCAGTTATTAGCGTGTCAGTGATGGTTGTGGCCTACCCTCCCCGAATCATAA
ACTACCTACCCAGGAACATGCTCAGGAGGACAGGGGAAGCCATGCAGCTCCACTGT
GTGGCCTTGGGAATCCCCAAGCCAAAAGTCACCTGGGAGACGCCAAGACACTCCCT
GCTCTCAAAAGCAACAGCAAGAAAACCCCATAGAAGTGAGATGCTTCACCCACAAG
GTACGCTGGTCATTCAGAATCTCCAAACCTCGGATTCCGGAGTCTATAAGTGCAGAG
CTCAGAACCTACTTGGGACTGATTACGCAACAACTTACATCCAGGTACTCTGACAGG
AAGGGGGAGACTAAAATTCAACAGAAGTCCACATCCACAGGGTTTATTTTTTGGAA
GAAGTTTAATCAAAGGCAGCCATAGGCATGTAAATGAGTCTGAATACATTTACAGT
ATTAAATTTACAATGGACATGCGATGAGACTTGTAAATGAAAGCATTGTGAACTGA
AACCGAGTCTCTGTGGATCTCAAAGCAAACTCTTAACTTAAGGCACTTTGATTTTGC
CAACAAATAATAACAAACATTAAGAGAAAAAAATGATCCACTACGAAATAACAAAC
GGCTAATGCACCTGAATTCTCAGTAAAAAGACCTTTCTCTCGCTAACAGTTGCCAGC
TGCCTCGTGTCTGTTTCCTACCAATGTCACAAACATCGCACACAGGGTGAATGGAGT
CAACGGGAAAGATTAAGTTTGCGGTCTGTGTAAATCTCAATGTACAAATATTCTGTC
NCTGGTTTATAAACATTTTGATAAAACCGAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAA (SEQ ID NO: 1)

Figure 2

SpeI (913)   EcoRI (938)

SpeI (249)

NotI (965)

Amp

EcoRI (1395)

CMV Promoter

608 flag in pcDNA3

XbaI (2175)

Neomycin

XbaI (3248)

13256bp

EcoRI (3266)

EcoRI (3384)

Flag

608

ApaI (8803)

EcoRI (5579)
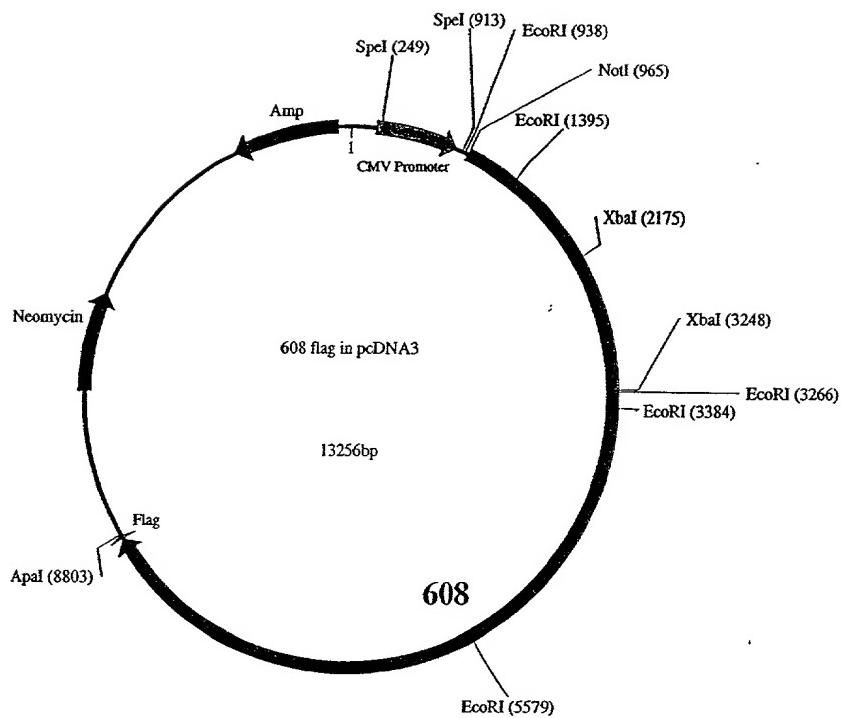
Figure 3

MQVRGREVSGLLISLTAVCLVVTPGSRACPRRCACYVPTEVHCTFRYLTSIPDGIPANVE
RINLGYNSLTRLTENDFDGLSKLELLMLHSNGIHRVSDKTFSGLQSLQVLKMSYNKVQII
RKDTFYGLGSLVRLHLDHNNIEFINPEAFYGLTSLRLVHLEGNRLTKLHPDTFVSLSYLQI
FKTSFIKYLFLSDNFLTSLPKEMVSYMPNLESLYLHGNPWTCDCHLKWLSEWMQGNPDI
IKCKKDRSSSSPQQCPLCMNPRISKGRPFAMVPSGAFLCTKPTIDPSLKSKSLVTQEDNGS
ASTSPQDFIEPFGSLSLNMTXXSGNKADMVCSIQKPSRTSPTAFTEENDYIMLNASFSTNL
VCSVDYNHIQPVWQLLALYSDSPLILERKPQLTETPSLSSRYKQVALRPEDIFTSIEADVR
ADPFWFQQEKIVLQLNRTATTLSTLQIQFSTDAQIALPRAEMRAERLKWTMILMMNNPK
LERTVLVGGTIALSCPGKGDPSPHLEWLLADGSKVRAPYVSEDGRILIDKNGKLELQMA
DSFDAGLYHCISTNDADADVLTYRITVVEPYGESTHDSGVQHTVVTGETLDLPCLSTGV
PDASISWILPGNTVFSQPSRDRQILNNGTLRILQVTPKDQGHYQCVAANPSGADFSSFKV
SVQKKGQRMVEHDREAGGSGLGEPNSSVSLKQPASLKLSASALTGSEAGKQVSGVHRK
NKHRDLIHRRRGDSTLRRFREHRRQLPLSARRIDPQRWAALLEKAKKNSVPKKQENTTV
KPVPLAVPLVELTDEEKDASGMIPPDEEFMVLKTKASGVPGRSPTADSGPVNHGFMTSI
ASGTEVSTVNPQTLQSEHLPDFKLFSVTNGTAVTKSMNPSIASKIEDTTNQNPIIIFPSVAE
IRDSAQAGRASSQSAHPVTGGNMATYGHTNTYSSFTSKASTVLQPINPTESYGPQIPITGV
SRPSSSDISSHTTADPSFSSHPSGSHTTASSLFHIPRNNNTGNFPLSRHLGRERTIWSRGRV
KNPHRTPVLRRHRHRTVRPAIKGPANKNVSQVPATEYPGMCHTCPSAEGLTVATAALS
VPSSSHSALPKTNNVGVIAEESTTVVKKPLLLFKDKQNVDIEIITTTTKYSGGESNHVIPTE
ASMTSAPTSVSLGKSPVDNSGHLSMPGTIQTGKDSVETTPLPSPLSTPSIPTSTKFSKRKTP
LHQIFVNNQKKEGMLKNPYQFGLQKNPAAKLPKIAPLLPTGQSSPSDSTTLLTSPPPALST
TMAATQNKGTEVVSGARSLSAGKKQPFTNSSPVLPSTISKRSNTLNFLSTETPTVTSPTAT
ASVIMSETQRTRSKEAKDQIKGPRKNRNNANTTPRQVSGYSAYSALTTADTPLAFSHSP
RQDDGGNVSAVAYHSTTSLLAITELFEKYTQTLGNTTALETTLLSKSQESTTVKRASDTP
PPLLSSGAPPVPTPSPPPFTKGVVTDSKVTSAFQMTSNRVVTIYESSRHNTDLQQPSAEAS
PNPEIITGTTDSPSNLFPSTSVPALRVDKPQNSKWKPSPWPEHKYQLKSYSETIEKGKRPA
VSMSPHLSLPEASTHASHWNTQKHAEKSVFDKKPGQNPTSKHLPYVSLPKTLLKKPRIIG
GKAASFTVPANSDVFLPCEAVGDPLPIIHWTRVSSGXEISQGTQKSRFHVLPNGTLSIQRV
SIQDRGQYLCSAFNPLGVDHFHVSLSVVFYPARILDRHVKEITVHFGSTVELKCRVEGMP
RPTVSWILANQTVVSETAKGSRKVWVTPDGTLIIYNLSLYDRGFYKCVASNPSGQDSLL
VKIQVITAPPVIIEQKRQAIVGVLGGSLKLPCTAKGTPQPSVHWVLYDGTELKPLQLTHS
RFFLYPNGTLYIRSIAPSVRGTYECIATSSSGSERRVVILTVEEGETIPRIETASQKWTEVN
LGEKLLLNCSATGDPKPRIIWRLPSKAVIDQWHRMGSRIHVYPNGSLVVGSVTEKDAGD
YLCVARNKMGDDLVLMHVRLRLTPAKIEQKQYFKKQVLHGKDFQVDCKASGSPVPEV
SWSLPDGTVLNNVAQADDSGYRTKRYTLFHNGTLYFNNVGMAEEGDYICSAQNTLGK
DEMKVHLTVLTAIPRIRQSYKTTMRLRAGETAVLDCEVTGEPKPNVFWLLPSNNVISFS
NDRFTFHANRTLSIHKVKPLDSGDYVCVAQNPSGDDTKTYKLDIVSKPPLINGLYANKT
VIKATAIRHSKKYFDCRADGIPSSQVTWIMPGNIFLPAPYFGSRVTVHPNGTLEMRNIRLS

DSADFTCVVRSEGGESVLVVQLEVLEMLRRPTFRNPFNEKVIAQAGKPVALNCSVDGNP
PPEITWILPDGTQFANRPHNSPYLMAGNGSLILYKATRNKSGKYRCAARNKVGYIEKLIL
LEIGQKPVILTYEPGMVKSVSGEPLSLHCVSDGIPKPNVKWTTPGGHVIDRPQVDGKYIL
HENGTLVIKATTAHDQGNYICRAQNSVGQAVISVSVMVVAYPPRIINYLPRNMLRRTGE
AMQLHCVALGIPKPKVTWETPRHSLLSKATARKPHRSEMLHPQGTLVIQNLQTSDSGVY
KCRAQNLLGTDYATTYIQVL (SEQ ID NO: 2)

Figure 4

Frag 1	Frag 2	Frag 3

75 KD →

50 KD →

35 KD →

30 KD →

25 KD →
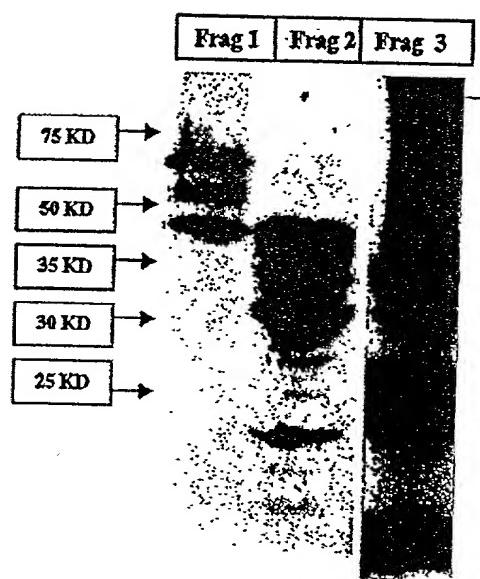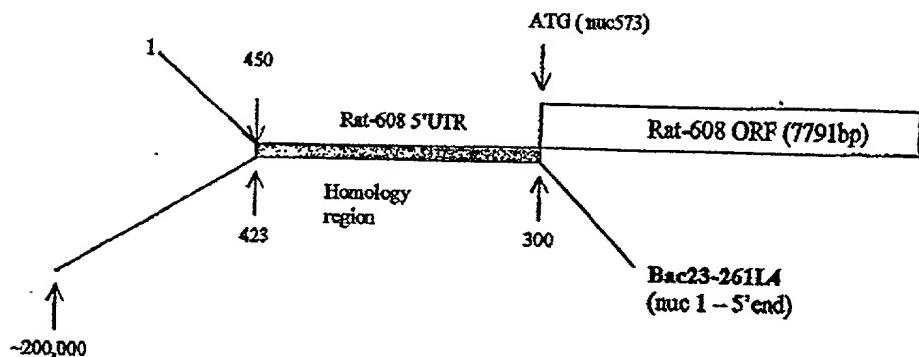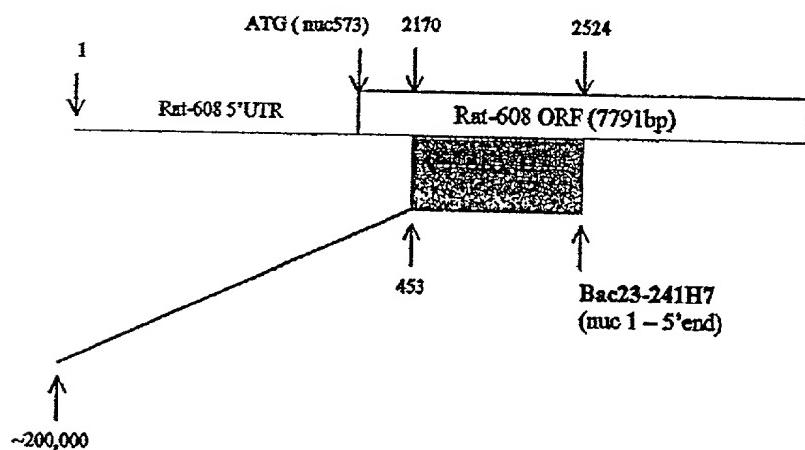
Figure 6

Figure 7

Nuc 1

↓

TTTGGAACCAACCCAGATGCCCCTCAACAGAGAAATGGGCCAGAAAATGTGGTCCATTTA
TCCAATGGAATACTACTCAACTTATTAAAAACAACGACTTTCATAAAATTTTTAGGCAAAT
GNATGGTCTGNAGGATCTTGAGTGAGGTAACCCAATCACAAAAGAACACTCATGGTATGC
ACTCACTGATAAGTGGCTATTTGTCTATGGAGTGATTTAAAAGGGAAGAAGACACATAGC
TTTTTGTGTGTATAATATTAAGATGGAAATTTGCCAGTGCTGTTTGGCTTATGAGTGAATCT
TGTTTCAGTGGATTACCGGAAGAAAATAATAAGTGAACTGTAGGAAGAAGTAGTTAATCA

↓ Nuc 390

AGGTGACAAAGTATCCTGACACATTGGGAAAAGACCACAGTCCAGGAAACTGAGTCTTAA
GGATTCATATTAACTCCAGTTCCCCATGTGCAGCTCTGAGACTTTGGCAGATCAGACACTT
AACTTCACCAGCTTCCTACACAGAGCAGTTACTATCCTTGCACTTCACAGATGGAGTGTGA
CCATTAAGCTGCACTGAAACATGAGTCTGACTTGTTAATAATCTTAAAAATACAAATTGTGT
TGTAAAGTATGTGACCAAAGAGCATGGTCATGCTATTAACCTTTGATGTTCTATGGACTCT
TAATTTTATGGTAGAAATGTCAACAAGCTTGTGGAGGCTGGAAGATACAAGGCTTAAGAG
GATGGCCTTTCAGTTTTGAAAGTAATTCAGTATGTGTTCTGGCATCCCTTTTCCTAAAGCA
ATTTAACCCCCCAAGTAGGCATAATTTTAATGCTTACTTCATCAGAATATGTCTAATTGAC
TCTTCTAAAAAGACTTTGGTATGCATAGGATCTAAATGTAAATGTGATTTACTGACATAAT
AAATAGGAGAAACTGAGCTAGAATAGGTATAAAATATGTGCTGGCTTTCTAATAGGTCTT
ATAGGTTATATAAGAGGTGGGAAAGGAATATTTGAAACATCTAGAAGTAAAATGATCCTG
AGTAGCGATCCTGGGAAAATACGTACTCTAACACACTGCAATCATCTCTCTGTGGTTTGCT
GGAGCTGAGGTCTGGAAGGCTCGACCTTGGTTAGAAATAACCTACCGAATACAGAGCTAT
GACGTTAGTCTGGAAGGGAGCTTTGGAAGAATGACAAGCTGTAGCTGCCCAGAACATACTA
GATGCCATATTTCCAAGGCAAGTGTCCACATGCGGACATCTTAAGAATATGGTTGTCTCTG
CAGTGCTAAGGACCTTGTTCGTGCCCACACAGGTCTCCAGGGTTAGTGCTAACTCTGACTGC
TTGACTCTTTAATTCTACCTTGATCATTAATGACTAGAAATCACTTGGTGATTAGCAACTG
GATATGGAATATTACTAATTTGTACCCAAGCCAGGCCACCTCAGCTTTGGCAGCTCCATTC
ATTCTGTGGAGCCCAGTCACGTGGGTTTGAATCAACTGTACTGTTTCTACTTACAAGACGC
ATTACCTGAGATGAGTCATTTTTCTTCACAAGTCTCTTTTAGAAGAGTCAATTAGACATATT
CTGATGAAGTAAGCATATAAAGTGAGAGCAGCATGAATGTGTTCCATGTATGCTCATGGA
TGCTATTATAATGTGGAAATAAACTGACTTTAAAAAAAAAAAGCTTATGATACTTGTCACA
GAGTAAATCTTCCATAAATATCATCTGCATTTATAAATTATTTTCATAATCCATCAATTAA
AAAACCTTTAGAAATTTTGTTAACACAAAGATCCCTAGGCCCCTGCCCTAGGATGGTCTGTA
TGGTGGGCCTGAGAGATGGAGCTTAAGAACTTACTTGCTCCAGGAGCACATCTTCAGAAC
ATCTGCCTCAAAACATTTATCCCAAATGCTCATCAAAAGGCTCACTCACATGTGCTTCAAACC
ACAGGGATTAAACAGTCATTTTAGTCACATTTCTCAAACGGTGGAAGCCTGCTAGAGGAA
CAGGATGTATCAGGATAACATCCAACCTTACAAAAGGATGTCATAACCCTCACCACAACA
AACAACAACGACAACAAACCCATAAAAATTATCACGGCAAATGAACTAAGCCATATGCA
GAAAAAGTATTATATGTTCTCATTGTGTGGGGTGTTTTTCCTTAATAGTCAAATATGCAGAAT
ATAGACAAAGATGGTTTATGCAAGTGGGGATGGCGAAGGATACTTGTAGATTAGAGGACA
CAAAGCAACAACTACAGAGTGAAGTAATCCAGAGACTTAATGTATAATATGAGGACTGTA
TTTAATAATTCTATTTAAGATACACAGCAAACGAGTGTATCTTACTAACACACACACTTAC
ATAGAGAGAATAAAGTGATAGATACGTTTGTTTTATCTTCATGTAGCTGATAATTTCATAT
TGTACACCTCAAACATAGATAACCAACAAAGAGGAAGAGGATAGGTGCCTCTCCCAGGGC
GGAAGAGTACATTCGAAAAGTCAGACACCATTGTGTAGATGTACCACATGGAGGAGCTAGA
GAAAGTAGCCAAGGAGCTAAAGGGATCTGCAACCCTATAGGTGGAACAACATTATGAGCT
AACCAGTACCCCGGAGCTCTTGACTCTAGCTGCATATATATCAAAAGATGGCCTAATCGG
CCATCACTGGAAAGAGAGGCCCATTGGACTTGCAAACTTTATATGCCCAGTACAGGGGA
ATACCAGGGCCAAAAAGGGGGAGTGGGTGGGCAGGGGAGTGGGGGTGGGTGGATATGGG
GGACTTTTGGTATAGCATTGGAAATGTAAATGAGTTAAATACCTAATAAAAAATGGAAAA

↑ Nuc 2817

AAAAAAAAAAAAAAAAAAAAAAAGGAAGGTCAGACACCTCACTTCACTGCTATCTCAACTT
GCAAACAGAAGGGGAGTCACAAACCCAGGACAAACCACAGTGATTGAAGCGTCTTTGAA
TGTTATTGCTGTTGTTGTTACCACCATCATTAGCATATATTCATTGTGAAAACTTACGGGGT
CTATGACATGTTTTTTTATTCAAGTATATCACATGCTGTCAGCATATTTGGCACCACTACCA

↓ Nuc 3114

GCCCCAGCCCCCTTTGCCCCGCCCCCAACACACACACACACACACAGACACACACACACA
CACACACACACACACACACACACACCTTTACCTTCTCCTGGGCATCATCTGCTCACTCACC
CACCCAAGCTTAATCCTTTTCCTTCCCTGCAATAGTACCTCTCCTATTTTTATGTCTAGGTT
CCCCCTCCCCCTGTTAGGAGATGGGAGAGGTCACGAAAGGAAAGAATTTGTAGCCCCTGA
GCCAGCCCGGGCCACAGAGCCTGCCACCAGACAGGAAAAGCCCAGGGCTTACCAGCACA
GGAGGAGCAAACTCGCAGGCGAGCCTGGGTTGGCGCTGGTGGTCCCGGGTCGATGGCCCG
CCCATTCCCAGAAGCCGAGGCTATAGCTGCGTCACCTGCCCCGCCCTCCTCCCGAGTGAAG
ACCCCTAGAGGCTGAGCAGACCCCAAAGGCGGTGCAATTCCATTGGCCCAAGGCAGAGGT
GAGCGGCTGCTAATCCCCTCGGGAAGTGAAGGGACCCAGAGAGAGTCTGGGTAGATGTGGGA

↓ Exon 1 (Nuc 3661)

GCTGGGGGTTCAGGGGCGAGACAGAGGGTGGGATGGGCAGAAGGGTCCAGGAAAAGGAAAGT
ACTGGAGGGGGAGTTGGGACAAAAGCAGCGACCAAGGGAACATCGCTTCAGTGACTGAAGCCA
GGCAAAAGGAGCGGGAAGGATTATATGTAGCCTGGGACGCTTTCATAAACACTGATGACGTGT
TTGTGCAAAGCAAGCAATTTGAGGAGAAACGCCTGGGACGTCGGAAAGAAGGAGTGATCGATT
AGTACTTGTAAGTTTAGGTGAGTTTGAGAACTAACTAACCTATACTATTGAGGGAGAAGGAAG
AGCATTCCAGCAGCAGCAGCAGCAGCAGCAATCAGATAAAGGAAAGCTTTGGTTAGTTTGGAA
ATGTATGATACCATTAAAATAACAGAAGCGCCTCCAGTTCTCTGAAGAGTCAGTCCCCCAGCTA
GTGAAGACTAAGCCTACTAAGCCTTTTGCTCCCGTTGGAAGCAAAGAACGTTCCTTCAATCAGG
TGAAGGCTCTCCTCAGAAGATTTCCTGTCTCTGCTTATGTTACAAGAGGATTCAAAAGCAAGAC
AGAAGAGCTCAGGTATTGCCAACTCTTT

↑ Nuc 4203

TGTTAAATACAGTTTGAGGCTTAAGTGTACGGGAACTCATGTGGTATTCATTTACGGCTCT
CTTCTCTTATAACTAACTCTTAAGGTGCATATAGTCTCTTCTGTTTCCCAGCTACCTTGTAC
CATCTTTGTTTATCTAATAATAGCAAGCTCATCTGCTTTTTAATCATCACGCAGAGAGTATT
CAAAAAATATTCAGTGATGTAACAGTGACAGTGTAGGCATAGAAGTAATCATTAGTAAATC
TTAATTTGGGTTAAACTCATTCATAACAGCTCCAGGTTGGGAGGGGATCACTGAGCCTTCGC
CACGTGCGGGTTAAAGATATTTTCTAACAAGAGAAGCAGAATTCTTCCTTGGCCATGCTCC
CCATCACTGTGTCAGTAAGCAGAGGGGTGTTTCCAAGCAGAGAAAAGAGCAGACAGTGTTA
TGCCTGCAAAGTCAGAGACTCAGCCCTCCCAGCTGGTCAGTTTACTGTCCTCCCGGTCATT
AGTTGGCTCTGAAAAGGCCCATGTGTCCTTATTGGCAAGGACTTGCAGACATGCTAGAAA
GAAATTTGACCTTTTTTTCTAGTGGGTTATTACAGCTGTAAAAGTATTTTGGAAGGGTTAAG
CCAAATAAATAAAACACATATTAAATAATACAATGTTACAAAAATTGATCATATAAAGAA
GTACATTCATAAATGCAATGTGAAAAAATATATATAATTTTTATCTATTTACTGGTGCAAAG
TTTTCTAAATTGCACATGTACTATTTTTATATTTATAAAAATATTTTTAAAAATGTATATAAA
AGTGTAAAAGGCTCTTGGTCAAACAAGAGAGTTAAAATTTACAAACTTTAATTGTCCCGAT
AACATTATTATGATCTCTAATGACAGGGATCCTGCTTTTCATTGGGAAATGAGGAAGCTATG
AAGATATGTTTACAATAATAAGCCCATTTAGTGATAAAGTCCAATGGGAAGCTAGCACAC
ACTGGTTTATAAAGAGAACAGTTTCCTGAGTCTATGCAAGTTTACACTCTAGGGAATAAG
AGTTCCTCTTTCTCCAGATTTCACTAGCATTGTTGTCATCATTTATCTTCTTGATGATGAG
CATTATAAGTGGAATAAGATAGGATCTCAAAGGAATGTCAATTTGGATGCCCTGAACAAT
CTTTCAGGTCTTTCTTTCAGTTCACTAGTCTATTCATTTATTGGATAATTGGGGGATGGTGT
TAATTTTTTTGCAGTTCTTATGGAATTCCAAAAAAACAAAAAACAAACAAACAAACAAAAA
ACCTCTGAAACTAGAACTACCAATCCATTACTGGGTATGTAACAAAGAGAAAATCTGCACA
GAATTTATTGCTACATTGTTCATTATTCACGACAGCCAAGAATGTGGAACCAACTTACGTA
GCCGTCAAAATATGAACGGATAAAAGAAAAATGTGGAAATGTGTACAACAGAGTCCCATGTG
GCCATAAAAGAGTGAAATCATGACATATGCAGGAAATGGATGCAACTGGAAATCAATTG
GGCTAATCAAAACAAGACAGACTCAAAAAGGAAACACCGTGTAGCTTCTCTGACAAACA
GAAGCTAGATTTACACTTGTACGTGCGCATGTGTGTTTAGAATTTTATTTAGTTATACACT
ATTCTAATCTGTGAGTGTGTATAAAGGCATGCATGTAAAGCAAAAACAAGCTAGCTGGGG
TGGGTAGGAGAGAAAGCAATGAGAGGGAGTTAATAAGAACGAAGCATAGTAACATAGGTG
CCAGGATGAAATGCATTAATTTGTATGCTAACTAAACCACAGACAGGAGGCACACGTTCA
AACCAGGGTGAAATCCCAGCACAGAGAAGGGGAAGTAGACACAAAGTTTCGCCACTAAC
CAAGAAGCCATTTGCAGTTGCTGCCTGCTGGGAGGGGCGTTCCAGTTTTCTCCAGTCTGAC
ACTGTGTATAACAACCAGTTGACAATACAAAGTTGGCATGATGGATGGTTTTTGTGCTATT
TTTCATTTTTTTTCTTACTGTTTTGTTGTTGTGGTGGTTGTTGTGGTGGTGGCTGTGGTTTTC
ATTTGTTTCTTTTGGAGAGGAGAGAGAGGAACATGAAATTGGGTGGGTAGGAAGCTGGAAACG
ATCTGGAAGAAGTTGGGGAAAGAGAAAAATTGTATGGAGCATATTTAAACAAACAAACA
AACAAACAAAAGGTTCATTTTGCCACAAAAAAGGTGTGAATTAAATTAACCAGTTACGACT
CTTAAAGAAAATATTCCCAATTATTCCCAGAGTTGCTATGTATGCTGTGCCTAGGACTTTG

▼ Exon 2 (Nuc 6559)

CTTGAACTGGCCCTATAACTCTCGGTGTGGTGTCTTTTCAGGATGCAGAAGAGAGGCAGGGAA
GTCAGCTGCTTGCTGATCTCCCTCACTGCCATCTGCCTGGTGGTCACCCCTGGGGAGCAGGGTC
TGTCCTCGCCGATGTGCCTGCTATGTGCCCACAGAGGTGCACTGTACATTTCGGGACCTGACC
TCCATCCCAGACGGGCATCCCAGCCAATGTGGAACGAGTCAATTTAGGGTGTGT

▲ Nuc 6755

GGACCTTGCCTGATCTCCTTCTCAGAGAGGGACCACTGATTTTCCTGGTACTTTGCCCCCC
AAACACCTGTGATTACTTTTAATAGTTTTCTTCTAAAATGGGTTCATACAAACCTTATATTG
TGGAGACAATGAACATTTTATCCCAATAGTCTTTTACTAGAACTTGAAGCCCCTCTTAGTT
GTTTGGGAGCCTCATAATTATGGGCAGCTTTATTCTGAATGAATTTTAAATGAAAAAGAT
ACAGTTTCTGTTAACAATCATTATGATACCAAGGAAGAGGAATTGTCATTGAATATTTTAA
AAAAGCATTTCTTTTGCAATTTATAAATACCCATTACAAAATGGCTTACTTAAAATACTTG
CCTTACTAAATCTGACAAATTATGGTGATATTTTGAAGGTTTATGAAAATTTGTTTATGTGT
ATAAATGCACAAGAAATGGGATATGCCATCACCTATGTGCCATTAGTGAGCATGTACAGT
ATGCCAAACACTATTGTTCACGTTTGGAGGAAGTAATGGGGGTGGGGGAGCAACAAGGGT
TATAACCGTATACCCAGTGCCTTGGAAGCGATTGCAAACAGTAAAGACTGACATTGTGTT
CTCCCTATGAGGGAGGGGCCTTGGGCTGAGCACTTTGCAATGAGCATTTGCTCATTGTGCT
GGCAGGTTTTATGATAACTTGACCCAAGCTAGAGTCACTGGAGAGGAAGGAACTTCAACT
GAGAACATGCCTGAAGAAGATCAGATTATAGGCAGGCCTGTGGGGCATTTTCTTAATTAG
TGATTCATGGGGCAGGGCCCAGTCCATTGTTCGTGGTACCATTTCTCAGGCACTATTAAAA
AAAAAAAAAACAGGCTGAGCAAGTGTCAAGGAGCAAGTCAGTGAGCAGCAGCCCTAATGA
TCTCTGCATCAGCTCCTGCCTCCAGGTTCCTACCCTATTTGAGTTCCTGTCCTAGCTCCCTA
CAGTGATGAACAATGATGTGGAAGTATAAGCCAAATAAATCCTTTCTTCCCCAACTTGCTG
TTGGTCATGATGTTTCATCACAGTGATAATAGTCCTCATGAAGATGCTGGTGTTTATAACA
CCTTTGGACTAAATTCTGTTATCTATAGCTGAGGAAAATGGAGCATAGAAAGTCTCCAGA
CTACACCAGAGTGTAATCTGGGCCTGAGCTTAGAATCACACCCACGTGCACTCCACTGCC
GGGGCTTCTTAACCGGAACACAGTTGTAAAAGGGAATTTTCTGTTTGTTTCCATTTTGACA

Exon 3 (Nuc 8089) ▼

TGTGGACTTTAATTGACGATTCATCTGAAGCTGAAAATGATTTTTTTTCCAGGTATAACAGC
CTCACTAGATTGACAGAAAATGACTTTTCTGGCCTGAGCAGACTGGAGTTACTCATGCTGCACA
GCAATGGCATTCACAGAGTCAGTGACAAGACCTTCTCGGGCTTGCAGTCCTTGCAG

Nuc 8218 ▲

GTGAGATAGGTAGAGGGTGATGGAGGCTGAGAAGAGAGGTGCAACTGTGGGTTATACCC
AAAAGCTGCTGATTCCCGTGGGAGACATTCTATAAGCATTCTATAAACTAGAGGCAGATA
TCAAGGAAGGATTTCAATTGTAATGCAATTTTATGAGAAAAATTTGAATATTAAGAAAATG
CTGGGGAAAATGCTTACACAATTGCGAGGACCTAATTTAGGATCTCCAATAGCCACATAA
AAAGCACAGCATGGCGGCAGACACCTGCAATTCCTGTCCCTGGAAGCACCTGTTCAGAAT
CCCAGAGACTCATTGGCCAAACACTCTATTCAATCAATGAAGTCCATATTCAGTGACAAA
ACTTGACTCAGAAACTAATGTGGAAAGCATCAGGAAGACAGCCAACATCTGGTCTCTACT
CATGCATGAATAAGGGATCCCAGAGAGAGAAGGGAAGAAAAAGGAAGGAAGGAAGGAAGG
AAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAGAGGGAGGAAAGGAGGGAGGG
AAGGAAGGAAAGGGAAAGGAAAAAAGAGATGGGGAGGGAGGGAAGGAAAGGAAAGGG
GGAGAAAGAAGAGAAGAAAGGAAAATAAATAAATTTTCAGGGATTATTACACCTTTAAA
TTTTATCCATAAAAGGTCATTTCCACCTGTTTGTCTGGAAGTAGAGTGGGATCCCTTATAT
AAGGGCAGTCTTTAACATAGTAGCATTTTATAAACCATTACAAATTTTGAGTTTTCTCTAC
TTTTTATCCTCTACCATCTTCAAACTGAAACTACAATTATTCCCACAAATGAAGAAAATGC
TGTAAGAGTTTTCACACACCGAAGTGGGAAACTTAAGGATTAGACAAGTCTAACAATGAG
AATGGGGAGAACAAAAAGAGACTGCACAGGGGAGCCCTTTCTCTGCTTATAATCTTGACAC
TTGAGAAGCTAATTGACGCTGCATGACTACTCAACTCTTTAAGCAAACAATGCTGTTGTTC
ATGAAAAGCACAATAAAGTACATATGTCCCATAATATTCATCAAAATTTGCATGCAGCAC
ATAATAGCAATCAAAGCAATAACACCCACTGTTCACAGAGACTTTAAACATGAAACTGGA
ACTATGTCTAGTGTTTTGACTTAGGGTACATAGTATGCTGTGTCTGTATGTACCAATGTTG
ATTTAGGTCATCAGACAGCATTTGGAACATGTATCTTCAGGAGGAATCATTCATGTATCCT
GCATGAAATTCTCCACCTATGTTTATTCTCTTAGCCAGGTTTTTCTCTGATGGAGAAACATT
GGGTTTGAGGTTTTACTCCCAGGTAACATTTAGGGAAAAGCTGTCTATGTTCTCAGTTTGG
CTTTTATTTATGAGGGATGTTGGTATTCCAGAAAATCTCTTTTGAAGAGATTACAATTTA
GGTCAAAACAGAAAAATATGTAAAAAGTTATTGTTTTTATTAGTATTTCATGTTCTTTTCTT
TTTTAAAAATGGTATGCTTAGAACTAATTAAGATTAGATTAGATTAGATTAGAAAATAATC
AGAGAGGGATTTGATGAATGCTAAAGCATCATGAAAAATTCAAAATTTTTTGCTTCTAATT
CAGAATCAATTAAATTCATATTACTATAAAAGACAGCACGCCAGATGTGTGCCAGCTGAG
GAGTGGATAAACTGTGTAACGTGAGTGCTATGTAGAAACAGAAAGGAGTGAAGGGTTGA

TGTGCGCTGCAACATCTTGAAAACATTCGGCTACATGATGGAAGCCAGGCACAAAAAGCC
ACATATTGCATGGTTATGTTTATATGAAATGTTTAAAATACATGGATTCTTAGCAAACAGA
GTAAGATGTTACTTAGGGTCAGGAAAAGATTAAAAAAAAAAAAAAACTATTGATGTGGAATG
ATCTTAATTTGGGGAAAAGACAATTTCCTAAGACGAAATAGTTGAGGTAGATATAGTTAT
ATCCCTGTGGATATTGTAATAAACCAGCATGCTGTGCTCTGAGAAGGGCCTAATGAAGGG
GCAGGAGGAAGTGAAATGAGATGGTAGAAAGGAAAGTCATATACCATGGCTTCTCTCGTG
GGTGGAATCTAGATATGTTAATATATTGACATAAAGGAAGGAATTGTTTAGGGAAGGATC
AAAACCAACAGGAGTGAGGGAGACAAATAGGAACCAATGAGAGGCAAAGTTCATGGTCAA
TGTGTGTGGAGACACCATAATAAAACTCCTTTTTTTGTTTGCTAACTAAAACCACTAAAATC
TAAAAACAAAACATTTTTGCACAAGAATTATTTATTATTCAATAAAGATGTTTAAATGGGG
GAAGTTGAAGTTCATTGATAGTCTCATAAATCTTAAATGTATTTAAACTGCTTTTTACGTTT
TTTATTATTAATTACTCTTGCTGTCATTATTATCATCATCATTATCGTCATCATCATCACTA
ATGCTTTTCACCATACACAAATGTAGGCAGAAGAGTGTAATCCACTTAGTGAGGCAATCTT
GGAGAGGGAAAGGAAGCGGATGCGGGGCAGAGGCCACACAGGAGGACAGTGAGAGGGAA
ATGAACAAGAAAAAATGTGGACACATGCACAAAAATTCCATAGTCCACTACATTACTTTG
TATTCTAATATTAAGAAAATAATAAACCCATTTCTGTGCACTTATCACCCAGGCTCAACAG
TTATCTTGGCCACAGATCCTGTCTCACTGCATCCTGTCCACCTGAGTCCACTTAGCGTTCTG
AATCCAATCCAGGGCATGATGCTTACTCCTACACAGAACTAAAGATTAAAGAGAGTTTAA
AAGTAACCATGACATCTCTCTGTTCCTTTAGCGATAAGTTCTTAATATTTATGGCTGCTTGT
GTATGTTCTAATTTCTCTAATATTGTCACATTTAGTTGGCAACTACTTTGTTTGAATTGAGT
TGGAGTTAAGGTCCCATAGGATTAATCTCAACATATTTCTATATTTATAAACTTTTCTCTCT

Exon 4 ? (Nuc 11286)↓

TTGTGAAAGTTCCTTTGAGAAAACAAATATGCCCATATCTTTCTTTACAGGTCTTAAAAATG
AGCTATAACAAAGTCCAAATAATTGAGAAGGATACTTTGTATGGACTCAGGAGCTTGACCCGG
TTGCACCTGGATCACAACAACATTGAGTTTATCAACCCCGAGGCGTTTTACGGACTCACCTTGC
TCCGCTTGGTACATCTAGAAGGAAACCGGCTGACAAAGCTCCATCCAGACACATTTGTCTCTTT
GAGCTATCTCCAGATATTTAAAACCTCCTTCATTAAGNACCTGTACTTGTATGATAACTTCATTG
ACCTCCCTCCCAAAAGAAATGGTCTCCTCTATGCCAAACCTAGAAAGCCTTTACTTGCATGGAA
ACCCATGGACCTGTGACTGCCATTTAAAGTGGTTGTCCGAGTGGATGCAGGGAAACCCAGGTA

Nuc 11680↑

ACTATCTTGTTTGTTTGTTTCTTTTTTTTATARKACGTATTTTCCTCAATTTCATTTAGAATGA
TATCCCAAAAGTCCCCCATAAACCTCCCCCCCACTTCCCTACCTACCCATTCCCATTTTTTGG
CCCTGGCATTCCCCTGTACTGGGGCATATAAAGTTTGCGTGTCCAATGGACCTCTCTTTCC
AGTGATGGCCAACTAGGCCATCTTTTGATACATATGCAGCTAGAGTCAAGAGCTCTGGGG
TACTGGTTAGTTCATAATGTTGTTGCACCTACAGGGTTGAA

Nuc 11967↑

(SEQ ID NO:3)

Figure 8

Exon/Intron No.	Exon start	Exon end	Exon length	Intron length
1	1	542	542	2356
2	2898	3094	197	1334
3	4428	4557	130	3068
4	7625	8019	394	

Figure 9
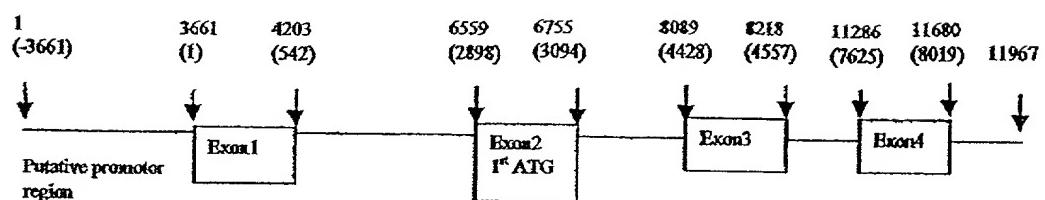
Figure 10

```
cDNA_rat       1    cgagagacgacagaaggttacggctgcgagaagacgacagaagggtccag
genomic_hu     1    --------------------------------------------------

cDNA_rat      51    aaaaaggaaagtgctggaggggagtggggacaaaagcagcgaccaagtga
genomic_hu     1    --------------------------------------------------

cDNA_rat     101    atgtcacttcagtgactgaggccaggcaaaacgcgcgggaaggattttgt
genomic_hu     1    --------------------------------------------------

cDNA_rat     151    gtagcttgggaccctttcatagacactgatgacacgtttacgcaaaatag
genomic_hu     1    --------------------------------------------------

cDNA_rat     201    aaatttgaggagaaacgcctgggccttcggaaaggagtgattgattagta
genomic_hu     1    --------------------------------------------------

cDNA_rat     251    cttgcaagtttaggtgactttaaggagaactaactaatgtatactattga
genomic_hu     1    --------------------------------------------------

cDNA_rat     301    gggaggaggaagagcattacagagtttccagcagcagcaggaaagctttg
genomic_hu     1    --------------------------------------------------

cDNA_rat     351    gttaatttggaaatggatgatagcattaaaataacagaagcgcctccagg
genomic_hu     1    --------------------------------------------------

cDNA_rat     401    totctgaagcttcagtcccccagctgaaagccagaaaagactaagccac
genomic_hu     1    --------------------------------------------------

cDNA_rat     451    taagccttttgatccctttggaagcaaagaactttccttccctggggtga
genomic_hu     1    --------------------------------------------------

cDNA_rat     501    agactctcctcagaagatttcctgtctctgcctatgttacaagaggaatc
genomic_hu     1    --------------------------------------------------

cDNA_rat     551    aaaaccaagacagaagagctcaggatgcaggtgagaggcagggaagtcag
genomic_hu     1    --------------------------------------------------

cDNA_rat     601    cggcttgttgatctccctcactgctgtctgcctggtggtcacccctggga
genomic_hu     1    --------------------------------------------------

cDNA_rat     651    gcagggcctgtcctcgccgctgtgcctgctatgtgcccacagaggtgcac
genomic_hu     1    --------------------------------------------------

cDNA_rat     701    tgtacatttcggtacctgacctccatcccagatggcatcccggccaatgt
genomic_hu     1    --------------------------------------------------

cDNA_rat     751    ggaacgaataaatttaggatataacagccttactagattgacagaaaacg
genomic_hu     1    --------------------------------------------------

cDNA_rat     801    actttgatggcctgagcaaactggagttactcatgctgcacagtaatggc
genomic_hu     1    --------------------------------------------------

cDNA_rat     851    attcacagagtcagtgacaagaccttctcgggcttgcagtccttgcaggt
genomic_hu     1    --------------------------------------------------

cDNA_rat     901    cttaaaaatgagctataacaaagtccaaatcattcggaaggatactttct
genomic_hu     1    --------------------------------------------------
```

```
cDNA_rat      951   acggactcgggagcttggtccggttgcacctggatcacaacaacattgaa
genomic_hu      1   --------------------------------------------------

cDNA_rat     1001   ttcatcaaccctgaggccttttatggacttacctcgctccgcttggtaca
genomic_hu      1   --------------------------------------------------

cDNA_rat     1051   tttagaaggaaaccggctcacaaagctccatccagacacatttgtctcat
genomic_hu      1   --------------------------------------------------

cDNA_rat     1101   taagctatctccagatatttaaaacctctttcattaagtacctgttcttg
genomic_hu      1   --------------------------------------------------

cDNA_rat     1151   tctgataacttcctgacctccctcccaaaagaaatggtctcctacatgcc
genomic_hu      1   --------------------------------------------------

cDNA_rat     1201   aaacctagaaagcctgtatttgcatggaaacccatggacctgtgactgcc
genomic_hu      1   --------------------------------------------------

cDNA_rat     1251   atttaaagtggttgtctgagtggatgcagggaaacccagatataataaaa
genomic_hu      1   --------------------------------------------------

cDNA_rat     1301   tgcaagaaagacagaagctcttccagtcctcagcaatgtcccctttgcat
genomic_hu      1   --------------------------------------------------

cDNA_rat     1351   gaaccccaggatctctaaaggcagacccttgctatggtaccatctggag
genomic_hu      1   --------------------------------------------------

cDNA_rat     1401   ctttcctatgtacaaagccaaccattgatccatcactgaagtcaaagagc
genomic_hu      1   --------------------------------------------------

cDNA_rat     1451   ctggttactcaggaggacaatggatctgcctccacctcacctcaagattt
genomic_hu      1   --------------------------------------------------

cDNA_rat     1501   catagaaccctttggctccttgtctttgaacatgacananntntctggaa
genomic_hu      1   --------------------------------------------------

cDNA_rat     1551   ataaggccgacatggtctgtagtatccaaaagccatcaaggacatcacca
genomic_hu      1   --------------------------------------------------

cDNA_rat     1601   actgcattcactgaagaaaatgactacatcatgctaaatgcgtcattttc
genomic_hu      1   --------------------------------------------------

cDNA_rat     1651   cacaaatcttgtgtgcagtgtagattataatcacatccagccagtgtggc
genomic_hu      1   --------------------------------------------------

cDNA_rat     1701   aacttctggctttatacagtgactctcctctgatactagaaaggaagccc
genomic_hu      1   --------------------------------------------------

cDNA_rat     1751   cagcttaccgagactccttcactgtcttctagatataaacaggtggctct
genomic_hu      1   --------------------------------------------------

cDNA_rat     1801   taggcctgaagacatttttaccagcatagaggctgatgtcagagcagacc
genomic_hu      1   --------------------------------------------------

cDNA_rat     1851   cttttggttccaacaagaaaaaattgtcttgcagctgaacagaactgcc
genomic_hu      1   --------------------------------------------------

cDNA_rat     1901   accacacttagcacattacagatccagttttccactgatgctcaaatcgc
genomic_hu      1   --------------------------------------------------

cDNA_rat     1951   tttaccaagggcggagatgagagcggagagactcaaatggaccatgatcc
genomic_hu      1   --------------------------------------------------
```

```
cDNA_rat      2001    tgatgatgaacaatcccaaactggaacgcactgtcctggttggcggcact
genomic_hu       1    --------------------------------------------------

cDNA_rat      2051    attgccctgagctgtccaggcaaaggcgacccttcacctcacttggaatg
genomic_hu       1    --------------------------------------------------

cDNA_rat      2101    gcttctagctgatgggagtaaagtgagagccccttacgttagcgaggatg
genomic_hu       1    --------------------------------------------------

cDNA_rat      2151    ggcgaatcctaatagacaaaaatgggaagttggaactgcagatggctgac
genomic_hu       1    --------------------------------------------------

cDNA_rat      2201    agctttgatgcaggtctttaccactgcataagcaccaatgatgcagatgc
genomic_hu       1    --------------------------------------------------

cDNA_rat      2251    ggatgttctcacatacaggataactgtggtagagccctatggagaaagca
genomic_hu       1    --------------------------------------------------

cDNA_rat      2301    cacatgacagtggagtccagcacacagtggttacggggtgagacgctcgac
genomic_hu       1    --------------------------------------------------

cDNA_rat      2351    cttccatgcctttccacgggtgttccagatgcttctattagctggattct
genomic_hu       1    --------------------------------------------------

cDNA_rat      2401    tccagggaacactgtgttctctcagccatcaagagacaggcaaattctta
genomic_hu       1    --------------------------------------------------

cDNA_rat      2451    acaatgggaccttaagaatattacaggttacgccaaaagatcaaggtcat
genomic_hu       1    --------------------------------------------------

cDNA_rat      2501    taccaatgtgtggctgccaaccoatcaggggccgacttttccagtttttaa
genomic_hu       1    --------------------------------------------------

cDNA_rat      2551    agtttcagttcaaaagaaaggccaaaggatggttgagcatgacagggagg
genomic_hu       1    --------------------------------------------------

cDNA_rat      2601    caggtggatctggacttggagaacccaactccagtgtttcccttaagcag
genomic_hu       1    --------------------------------------------------

cDNA_rat      2651    ccagcatctttgaaactctctgcatcagctttgacagggtcagaggctgg
genomic_hu       1    --------------------------------------------------

cDNA_rat      2701    aaaacaagtctccggtgtacataggaagaacaaacatagagacttaatac
genomic_hu       1    --------------------------------------------------

cDNA_rat      2751    atcggcggcgtggggattccacgctccggcgattcagggagcataggagg
genomic_hu       1    --------------------------------------------------

cDNA_rat      2801    cagctccctctctctgctcggagaattgacccgcaacgctgggcagcact
genomic_hu       1    --------------------------------------------------

cDNA_rat      2851    tctagaaaaagccaaaaagaattctgtgccaaaaaagcaagaaaatacca
genomic_hu       1    --------------------------------------------------

cDNA_rat      2901    cagtaaagccagtgccactggctgttcccctcgtggaactcactgacgag
genomic_hu       1    --------------------------------------------------

cDNA_rat      2951    gaaaaggatgcctctggcatgattcctccagatgaagaattcatggttct
genomic_hu       1    --------------------------------------------------
```

```
cDNA_rat      3001   gaaaactaaggcttctggtgtcccaggaaggtcaccaactgctgactctg
genomic_hu       1   --------------------------------------------------

cDNA_rat      3051   gaccagtaaatcatggtttttatgacgagtatagcttctggcacagaagtc
genomic_hu       1   --------------------------------------------------

cDNA_rat      3101   tccactgtgaatccacaaacactacaatctgagcaccttcctgatttcaa
genomic_hu       1   --------------------------------------------------

cDNA_rat      3151   attatttagtgtaacaaacggtacagctgtgacaaagagtatgaacccat
genomic_hu       1   --------------------------------------------------

cDNA_rat      3201   ccatagcaagcaaaatagaagatacaaccaaccaaaacccaatcattatc
genomic_hu       1   --------------------------------------------------

cDNA_rat      3251   tttccatcagtagctgaaattcgagattctgctcaggcaggaagagcatc
genomic_hu       1   --------------------------------------------------

cDNA_rat      3301   ttcccaaagtgcacacoctgtaacaggggggaaacatggctacctatggcc
genomic_hu       1   --------------------------------------------------

cDNA_rat      3351   ataccaacacatatagtagctttaccagcaaagccagtacagtcttgcag
genomic_hu       1   --------------------------------------------------

cDNA_rat      3401   ccaataaatccaacagaaagttatggacctcagatacctattacaggagt
genomic_hu       1   --------------------------------------------------

cDNA_rat      3451   cagcagacctagcagtagtgacatctcttctcacactactgcagaccota
genomic_hu       1   --------------------------------------------------

cDNA_rat      3501   gcttctccagtcacccttcaggttcacacaccactgcctcgtctttattt
genomic_hu       1   --------------------------------------------------

cDNA_rat      3551   cacattcctagaaacaacaatacaggtaacttcccccttgtccaggcactt
genomic_hu       1   --------------------------------------------------

cDNA_rat      3601   gggaagagagaggacaatttgggagcagagggagagttaaaaacccacata
genomic_hu       1   --------------------------------------------------

cDNA_rat      3651   gaaccccagttctccgacggcatagacacaggactgtgaggccagcaatc
genomic_hu       1   --------------------------------------------------

cDNA_rat      3701   aagggacctgctaacaaaaatgtgagccaagttccagccacagagtacca
genomic_hu       1   --------------------------------------------------

cDNA_rat      3751   tgggatgtgccacacatgtccttccgcagaggggctcacagtggctactg
genomic_hu       1   --------------------------------------------------

cDNA_rat      3801   cagcactgtcagttccaagttcatcccacagtgccctccccaaaactaat
genomic_hu       1   --------------------------------------------------

cDNA_rat      3851   aatgttggggtcatagcagaagagtctaccactgtggtcaagaaaccact
genomic_hu       1   --------------------------------------------------

cDNA_rat      3901   gttactatttaaggacaaacaaaatgtagatattgagataataacaacca
genomic_hu       1   --------------------------------------------------

cDNA_rat      3951   ctacaaaatattccggaggggaaagtaaccacgtgattcctacggaagca
genomic_hu       1   --------------------------------------------------

cDNA_rat      4001   agcatgacttctgctccaacatctgtatccctggggaaatctcctgtaga
genomic_hu       1   --------------------------------------------------

cDNA_rat      4051   caatagtggtcacctgagcatgcctgggaccatccaaactgggaaagatt
genomic_hu       1   --------------------------------------------------

cDNA_rat      4101   cagtggaaacaacaccacttcccagccccctcagcacaccctcaataccа
genomic_hu       1   --------------------------------------------------
```

```
cDNA_rat      4151   acaagcacaaaattctcaaagaggaaaactccottgcaccagatctttgt
genomic_hu       1   --------------------------------------------------

cDNA_rat      4201   aaataaccagaagaaggagggatgttaaagaatccatatcaattcggtt
genomic_hu       1   --------------------------------------------------

cDNA_rat      4251   tacaaaagaacccagccgcaaagcttcccaaaatagctcctcttttaccc
genomic_hu       1   --------------------------------------------------

cDNA_rat      4301   acaggtcagagttccccctcagattctacaactctcttgacaagtccgcc
genomic_hu       1   --------------------------------------------------

cDNA_rat      4351   accagctctgtctacaacaatggctgccactcagaacaagggcactgaag
genomic_hu       1   --------------------------------------------------

cDNA_rat      4401   tagtatcaggtgccagaagtctctcagcagggaagaagcagcccttcacc
genomic_hu       1   --------------------------------------------------

cDNA_rat      4451   aactcctctccagtgcttcctagcaccataagcaagagatctaatacatt
genomic_hu       1   --------------------------------------------------

cDNA_rat      4501   aaacttcttgtcaacggaaacccccacagtgacaagtcctactgctactg
genomic_hu       1   --------------------------------------------------

cDNA_rat      4551   catctgtcattatgtctgaaacccaacgaacaagatccaaagaagcaaaa
genomic_hu       1   --------------------------------------------------

cDNA_rat      4601   gaccaaataaaggggcctcggaagaacagaaacaacgcaaacaccacccc
genomic_hu       1   --------------------------------------------------

cDNA_rat      4651   caggcaggtttctggctatagtgcatactcagctctaacaacagctgata
genomic_hu       1   --------------------------------------------------

cDNA_rat      4701   cccccttggctttcagtcattccccacgacaagatgatggtggaaatgta
genomic_hu       1   --------------------------------------------------

cDNA_rat      4751   agtgcagttgcttatcactcaacaacctctcttctggccataactgaact
genomic_hu       1   --------------------------------------------------

cDNA_rat      4801   gtttgagaagtacacccagactttgggaaatacaacagctttggaaacaa
genomic_hu       1   --------------------------------------------------

cDNA_rat      4851   cgttgttgagcaaatcacaggagagtaccacagtgaaaagagcctcagac
genomic_hu       1   --------------------------------------------------

cDNA_rat      4901   acaccaccaccactcctcagcagtggggcgcccccagtgcccactccttc
genomic_hu       1   --------------------------------------------------

cDNA_rat      4951   cccacctccttttactaagggtgtggttacagacagcaaagtcacatcag
genomic_hu       1   --------------------------------------------------

cDNA_rat      5001   ctttccagatgacgtcaaatagagtggtcaccatatatgaatcttcaagg
genomic_hu       1   --------------------------------------------------

cDNA_rat      5051   cacaatacagatctgcagcaaccctcagcagaggctagccccaatcctga
genomic_hu       1   --------------------------------------------------

cDNA_rat      5101   gatcataactggaaccactgactctccctctaatctgtttccatccactt
genomic_hu       1   --------------------------------------------------

cDNA_rat      5151   ctgtgccagcactaagggtagataaaccacagaattctaaatggaagccc
genomic_hu       1   --------------------------------------------------

cDNA_rat      5201   tctccctggccagaacacaaatatcagctcaagtcatactccgaaaccat
genomic_hu       1   --------------------------------------------------
```

```
cDNA_rat      5251  tgagaagggcaaaaggccagcagtaagcatgtccccccacctcagccttc
genomic_hu       1  --------------------------------------------------

cDNA_rat      5301  cagaggccagcactcatgcctcacactggaatacacagaagcatgcagaa
genomic_hu       1  --------------------------------------------------

cDNA_rat      5351  aagagtgtttttgataagaaacctggtcaaaacccaacttccaaacatct
genomic_hu       1  --------------------------------------------------

cDNA_rat      5401  gccttacgtctctctacctaagactctattgaaaaagccaagaataattg
genomic_hu       1  --------------------------------------------------

cDNA_rat      5451  gaggaaaggctgcaagctttacagttccagctaattcagacgttttttctt
genomic_hu       1  --------------------------------------------------

cDNA_rat      5501  ccttgtgaggctgttggagacccactgcccatcatccactggaccagagt
genomic_hu       1  --------------------------------------------------

cDNA_rat      5551  ttcatcagganttgaaatatcccaagggacacagaaaagccggttccacg
genomic_hu       1  --------------------------------------------------

cDNA_rat      5601  tgcttcccaatggcaccttgtccatccagagggtcagtattcaggaccgt
genomic_hu       1  --------------------------------------------------

cDNA_rat      5651  ggacagtacctgtgctctgcatttaatccactgggcgtagaccatttttca
genomic_hu       1  --------------------------------------------------

cDNA_rat      5701  tgtctctttgtctgtggtttttttaccccggcaaggattttggacagacatg
genomic_hu       1  --------------------------------------------------

cDNA_rat      5751  tcaaggagaccacagttcactttggaagtactgtggaactaaagtgcaga
genomic_hu       1  --------------------------------------------------

cDNA_rat      5801  gtggagggtatgccgaggcctacggtttcctggatacttgcaaaccaaac
genomic_hu       1  --------------------------------------------------

cDNA_rat      5851  ggtggtctcagaaacggccaagggaagcagaaaggtctgggtaacacctg
genomic_hu       1  --------------------------------------------------

cDNA_rat      5901  atggaacattgatcatctataatctgagtctttatgatcgtggttttttac
genomic_hu       1  --------------------------------------------------

cDNA_rat      5951  aagtgtgtggccagcaacccatctggccaggattcactgttggttaagat
genomic_hu       1  --------------------------------------------------

cDNA_rat      6001  acaagtcatcacagctcccctgtcattatagagcaaaagaggcaagcca
genomic_hu       1  --------------------------------------------------

cDNA_rat      6051  tcgttgggggtttttaggtggaagtttgaaactgccctgcactgcaaaagga
genomic_hu       1  --------------------------------------------------

cDNA_rat      6101  actccccagcctagtgttcactgggtcctttatgatgggactgaactaaa
genomic_hu       1  --------------------------------------------------

cDNA_rat      6151  accattgcagttgactcattccagattttttcttgtatccaaatggaactc
genomic_hu       1  --------------------------------------------------

cDNA_rat      6201  tgtatataagaagcatcgctccttcagtgaggggcacttatgagtgcatt
genomic_hu       1  --------------------------------------------------

cDNA_rat      6251  gccaccagctcctcaggctcagagagaagggtagtgattcttactgtgga
genomic_hu       1  --------------------------------------------------
```

```
cDNA_rat     6301    agagggagagacaatccccaggatagaaactgcctctcagaaatggactg
genomic_hu      1    --------------------------------------------------

cDNA_rat     6351    aggtgaatttgggtgagaaattactactgaactgctcagctactggggat
genomic_hu      1    --------------------------------------------------

cDNA_rat     6401    ccaaagcctagaataatctggaggctgccatccaaggctgtcatcgacca
genomic_hu      1    --------------------------------------------------

cDNA_rat     6451    gtggcacagaaTGGGCAGCCGAATCCACGTCTACCCAAATGGATCCTTGg
genomic_hu      1    -----------TGGGCAGCTGGATCCACGTCTACCCTAATGGATCCCTGt

cDNA_rat     6501    tggttgggTCAGTGACGGAAAAAGACGCTGGTGACTACTTATGTGTGGCA
genomic_hu     40    ttattggaTCAGTAACAGAAAAAGACAGTGGTGTCTACTTGTGTGTGGCA

cDNA_rat     6551    AGAAACAAAATGGGAGATGACCTAGTCCTGATGCATGTccgcctgAGATT
genomic_hu     90    AGAAACAAAATGGGGGATGATCTGATACTGATGCATGTtagcctaAGACT

cDNA_rat     6601    GACACCTGCCAAAATTGAACAGAAGCAGTATTTTAagAAGCAAGTGCTCC
genomic_hu    140    GAAACCTGCCAAAATTGACCACAAGCAGTATTTTAgaAAGCAAGTGCTCC

cDNA_rat     6651    ATGGGAAAGATTTCCAAGTTGACTGCAaggcctctGGCTCCCCTGTGCCT
genomic_hu    190    ATGGGAAAGATTTCCAAGTAGATTGCAaagcttccGGCTCCCCAGTGCCA

cDNA_rat     6701    GAGGTATCCTGGAGTTTGCCTGATGggacagtgcTCAACAATGTAGCCCA
genomic_hu    240    GAGATATCTTGGAGTTTGCCTGATGgaaccatgaTCAACAATGCAATGCA

cDNA_rat     6751    AGCTGATGACAGTGGCTATAGGACcaagaggtacACCCTTTTCCACAATG
genomic_hu    290    AGCCGATGACAGTGGCCACAGGACtaggagatatACCCTTTTCAACAATG

cDNA_rat     6801    GAACCTTGTATTTCAACAACGTTGggatggcaGAGGAAGGAGATTATATC
genomic_hu    340    GAACTTTATACTTCAACAAAGTTGgggtagcgGAGGAAGGAGATTATACT

cDNA_rat     6851    TGCTCTGCCCAGAACACCTTAGGGAAAGATGAGATGAAAGTCCACCTAAC
genomic_hu    390    TGCTATGCCCAGAACACCCTAGGGAAAGATGAAATGAAGGTCCACTTAAC

cDNA_rat     6901    AGTTCTAACAGCcatcCCACGGATAAGGCAAAGCTACAAGACCACCATGA
genomic_hu    440    AGTTATAACAGCtgctCCCCGGATAAGGCAGAGTAACAAAACCAACAAGA

cDNA_rat     6951    GGCTCAggGCTGGAGAAACAGCTGTCCTTGACTGCGAGGTCACTGGGGaa
genomic_hu    490    GAATCAaaGCTGGAGACACAGCTGTCCTTGACTGTGAGGTCACTGGGGat

cDNA_rat     7001    ccgaagcccaatgTATTTTGGTTGCTGCCTTCCAAcaatgtcATTTCATT
genomic_hu    540    cccaaccoaaaaaTATTTTGGTTGCTGCCTTCCAAtgacatgATTTCCTT

cDNA_rat     7051    CTCCAATGACAGGTTCACATTTCATGCCAATagaaCTTTGTCCATCCATA
genomic_hu    590    CTCCATTGATAGGTACACATTTCATGCCAATgggtCTTTGACCATCAACA

cDNA_rat     7101    AAGTGAAACCACTTGACTCTGGGGActatgtgtgcgtagctcagAATCCT
genomic_hu    640    AAGTGAAACTGCTCGATTCTGGAGtacgtatgtgtagcccgaAATCCC

cDNA_rat     7151    AGTGGGGATGACACTAAGACATACAAACTGGAcattGTCTCTAAACCTCC
genomic_hu    690    AGTGGGGATGACACCAAAATGTACAAACTGGAtgtgGTCTCTAAACCTCC

cDNA_rat     7201    ATTAATCAATGGCCTGTATGCAAACAagACTGTTATTAAAGCCACAGCCa
genomic_hu    740    ATTAATCAATGGTCTGTATACAAACAgaACTGTTATTAAAGCCACAGCtg

cDNA_rat     7251    ttcggcacTCCAAAAAATACTTTGACTGCAGAGCAGATGGGATCCCATct
genomic_hu    790    tgagacatTCCAAAAAACACTTTGACTGCAGAGCTGAAGGGACACCATct

cDNA_rat     7301    tcccaggtcacGTGGATTATGCCAGGCAATATTTTCCTCCCAGCTCCATA
genomic_hu    840    cctgaagtcatGTGGATCATGCCAGACAATATTTTTCCTCACAGCCCCATA
```

```
cDNA_rat      7351   CttTGGAAGCAGAGTCACGGTCCATCCAAATGGAACCTTGGAgatgagga
genomic_hu     890   CtaTGGAAGCAGAATCACAGTCCATAAAAATGGAACCTTGGAaattagga

cDNA_rat      7401   acatccGGCTTTCTGACTCTGCGGACTTCACCTGTGTGGttcggagcgag
genomic_hu     940   atgtgaGGCTTTCAGATTCAGCCGACTTTATCTGTGTGGcccgaaatgaa

cDNA_rat      7451   ggaGGAGAGAGTGTGTTGGTAGTGCAGTTAGAAGTCCTAGAAATGCTGAG
genomic_hu     990   ggtGGAGAGAGCGTGTTGGTAGTACAGTTAGAAGTACTGGAARTGCTGAG

cDNA_rat      7501   AAGACCAACATTCAGAAACCCATTCAACGAAAAagtcatcgcccaagctg
genomic_hu    1040   AAGACCGACATTTAGAAATCCATTTAATGAAAAaatagttgcccagctgg

cDNA_rat      7551   gcaagcccgtAGCACTGAACTGCTCTGTGGATGGGAACCCCCCACCTGAA
genomic_hu    1090   gaaagtccacAGCATTGAATTGCTCTGTTGATGGTAACCCACCACCTGAA

cDNA_rat      7601   attacCTGGATCTTACCTGACGGCACACAGTTTGCTAACAGACCAcacaa
genomic_hu    1140   ataatCTGGATTTTACCAAATGGCACACGATTTTCCAATGGACCAcaaag

cDNA_rat      7651   ttccccGTATCTGATGGCAGGCAATGGCTCTCTCATCCTTTACAAAgcaa
genomic_hu    1190   ttatcaGTATCTGATAGCAAGCAATGGTTCTTTTATCATTTCTAAAacaa

cDNA_rat      7701   ctcggaacaagtCAGGGAAGTATCGCTGTGCAGCCAGGAATAAGGTTGGC
genomic_hu    1240   ctcgggaggatgCAGGAAAATATCGCTGTGCAGCTAGGAATAAAGTTGGC

cDNA_rat      7751   TAcatcgagaaactcatcctgttagagATTGGGCAGAAGCCAGTCATTCT
genomic_hu    1290   TAtattgagaaattagtcatattagaaATTGGCCAGAAGCCAGTTATTCT

cDNA_rat      7801   GACATACGAACCAGGGAtggtgaagagcgTCAGTGGGGAACCGTTATCAC
genomic_hu    1340   TACCTATGCACCAGGGAcagtaaaaggcaTCAGTGGAGAATCTCTATCAC

cDNA_rat      7851   TGCATTGTGTGTCTGATGGgATCCCCAAGCCAAATGTCAAGTGGACTACA
genomic_hu    1390   TGCATTGTGTGTCTGATGGaATCCCTAAGCCAAATATCAAATGGACTATG

cDNA_rat      7901   CCGGGTGGccATGTAATCGACAGGCCTCAAGTGGATGGAAAATACATAcT
genomic_hu    1440   CCAAGTGGttATGTAGTAGACAGGCCTCAAATTAATGGGAAATACATAtT

cDNA_rat      7951   GCATGAAAATGGCACGCTGGTCATCAAAGCAACAACAGCTcacgaccaAG
genomic_hu    1490   GCATGACAATGGCACCTTAGTCATTAAAGAAGCAACAGCTttatgacagAG

cDNA_rat      8001   GAAATTATATCTGTAGGGCTCAAAACAGTGTTGGCCAggcagttattagc
genomic_hu    1540   GAAACTATATCTGTAAGGCTCAAAATAGTGTTGGTCAtacactgattact

cDNA_rat      8051   gtgtCAGTGATGGTTGTGGCCTACCCTCCCCGAATCATAAActacctACC
genomic_hu    1590   gttcCAGTAATGATTGTAGCCTACCCTCCCCGAATTACAAAtcgtccACC

cDNA_rat      8101   CAGGAACATGCTCAGGAGGACAGGGGAAGCCatgCAGCTCCACTGTGTGG
genomic_hu    1640   CAGGAGTATTGTCACCAGGACAGGGGCAGCCtttCAGCTCCACTGTGTGG

cDNA_rat      8151   CCTTGGGAATCCCCAAGCCAaAAGTCACCTGGGAGACGCCAAGACACTCC
genomic_hu    1690   CCTTGGGAGTTCCCAAGCCAgAAATCACATGGGAGATGCCTGACCACTCC

cDNA_rat      8201   CTGCTCTCAAaagcaacagcaagaaaaccCCATAGAAGTGAGATGCTTCA
genomic_hu    1740   CTTCTCTCAAcggcaagtaaagagaggacaCATGGAAGTGAGCAGCTTCA

cDNA_rat      8251   CCCACAAGGTACGCTgGTCATTCAGAATCTCCAAACCTCGGATTCCGGag
genomic_hu    1790   CTTACAAGGTACCCTaGTCATTCAGAATCCCCAAACCTCCGATTCTGGga

cDNA_rat      8301   tcTATAAGTGCAGAGCTCAGAACCTACTTGGgacTGATTACGCAACAACT
genomic_hu    1840   taTACAAATGCACAGCAAAGAACCCACTTGGtagTGATTATGCAGCAACG

cDNA_rat      8351   TACATCCAGGTACTCTGACAGGAAgggggagactaaaattcaacagaagt
genomic_hu    1890   TATATTCAAGTAATCTGACATGAAataataaagtcaacaacatctgggca
```

```
cDNA_rat    8401   ccacatccacagggTTTATTTTTTGGAAGAAGTTTAATCAAAGGCAGCCA
genomic_hu  1940   gaa-----------TTTATTTTTTGGAAGAAGTTTAATCAAAGGCAGCCA

cDNA_rat    8451   TAGGCATGTAAATGAgtcTGAATACATTTACAGTATTAAATTTACAATGG
genomic_hu  1979   TAGGCATGTAAATGAattTGAATACATTTACAGTATTAAATTTACAATGA

cDNA_rat    8501   ACATGCgatga-----GACTTGTAAATGAAAGCATTGTGAACTGAaaccg
genomic_hu  2029   ACATGCaaaataaaagGACTTGTAAATAAATGCATTATGAACTGAtgata

cDNA_rat    8546   agtctctg-----TGGATCTCAAAGCAAACTCTTAACTTAAGGCACTTTg
genomic_hu  2079   ctgatttatttaaTGGATCTCAAAACAAACTTTTAACTTAAGGCACTTTt

cDNA_rat    8591   ATTTTGCCAACAAATAATAACAAACAttaagagaaaaaaatgatcCACTA
genomic_hu  2129   ATTTTGCCAACAAATAACAATAAACAaacattgaaacggtt----CACTA

cDNA_rat    8641   CGAAATAACAAACGGCTAATGCACCTGAATTctcagtaaaaagacctttc
genomic_hu  2175   TAAAATAACAAATGGCTAATGTACCTGAATTtttcagtaaaaaaatgaac

cDNA_rat    8691   tctcgctaacagttgCCAGCTGCCTCGTGTCTGTTTCCTACCAATGTCAC
genomic_hu  2225   ttctaata-------CCAGTTGCCTAGTGTCCACCTCCTATCAATGTTAC

cDNA_rat    8741   AAACAtcgcacacagggtGAATGGAGTCAACGGGAAAGATTAAGTTTGCG
genomic_hu  2268   AAGCAtggcactca----GAACAGAGACAATGGAAAATATTAAATCTGCA

cDNA_rat    8791   GTCtgtgtaaatctca-----------ATGTACAAATATTCTGtcncTG
genomic_hu  2314   ATCtttatgatgtaaatttaccatcctgATGTATAAATATTTTG----TG
```

exon1 (2342-2397)

```
cDNA_rat    8791   GTCtgtgtaaatctca-----------ATGTACAAATATTCTGtcncTG
genomic_hu  2314   ATCtttatgatgtaaatttaccatcctgATGTATAAATATTTTG----TG


cDNA_rat    8829   GTTTATAAACATTTTGATAAAACCGAAAAAAAAAAAAAAaaaaaaaaaaaa
genomic_hu  2360   GTTTATAAATTTTTTTGCTAAAACCTACAGAAAATAAGcactgaactgtc
```

(Genomic human OCP: SEQ ID NO:4)
(cDNA rat: SEQ ID NO:5)

Figure 11

Exon/Intron No.	Exon start	Exon end	Exon length	Intron length	Remarks
1	1	208	208	69	No valid splice site found upstream this exon
2	277	429	153	18	
3	447	485	39	1561	
4	2046	2244	199	1351	
5	3595	3724	130	3254	
6	6978	7359	382	4123	
7	11482	14903	3422	38	
8	14941	15307	367	51	
9	15358	15825	468	1039	
10	16864	~17760	~897	?	Last exon might be complete
11	?(1)	2317	~2317	25	Exon is not complete and start site is not known
12	2342	2397	56	-	

FIGURE 12

608 Human translated nucleotide sequence (ORF)
ATGAAGGTAAAAGGCAGAGGAATCACCTGCTTGCTGGTCTCCTTTGCTGT
GATCTGCCTGGTCGCCACCCCTGGGGGCAAGGCCTGTCCTCGCCGCTGTG
CCTGTTATATGCCTACGGAGGTACACTGCACATTTCGGTACCTGACTTCCA
TCCCAGACAGCATCCCGCCCAATGTGGAACGCATCAATTTAGGATACAAC
AGCTTGGTTAGATTGATGGAAACAGATTTTTCTGGCCTGACCAAACTGGA
GTTACTCATGCTTCACAGCAATGGCATTCACACAATCCCTGACAAGACCTT
CTCAGATTTGCAGGCCTTGCAGGTCTTAAAAATGAGCTATAATAAAGTCC
GAAAACTTCAGAAAGATACTTTTTATGGCCTCAGGAGCTTGACACGATTG
CACATGGACCACAACAATATTGAGTTTATAAACCCAGAGGTTTTTTATGG
GCTCAACTTTCTCCGCCTGGTGCACTTGGAAGGAAATCAGCTCACTAAGCT
CCACCCAGATACATTTGTCTCTTTGAGCTACCTCCAGATATTTAAAATCTC
TTTCATTAAGTTCCTATACTTGTCTGATAACTTCCTGACCTCCCTCCCTCAA
GAGATGGTCTCCTATATGCCTGACCTAGACAGCCTTTACCTGCATGGAAA
CCCATGGACCTGTGATTGCCATTTAAAGTGGTTGTCTGACTGGATACAGCC
AGATGTAATAAAATGCAAAAAAGATAGAAGTCCCTCTAGTGCTCAGCAGT
GTCCACTTTGCATGAACCCTAGGACTTCTAAAGGCAAGCCGTTAGCTATG
GTCTCAGCTGCAGCTTTCCAGTGTGCCAAGCCAACCATTGACTCATCCCTG
AAATCAAAGAGCCTGACTATTCTGGAAGACAGTAGTTCTGCTTTCATCTCT
CCCCAAGGTTTCATGGCACCCTTTGGCTCCCTCACTTTGAATATGACAGAT
CAGTCTGGAAATGAAGCTAACATGGTCTGCAGTATTCAAAAGCCCTCAAG
GACATCACCCATTGCATTCACTGAAGAAAATGACTACATCGTGCTAAATA
CTTCATTTTCAACATTTTTGGTGTGCAACATAGATTACGGTCACATTCAGC
CAGTGTGGCAAATTTTGGCTTTGTACAGTGATTCTCCTCTGATACTAGAAA
GGAGCCACTTGCTTAGTGAAACACCGCAGCTCTATTACAAATATAAACAG
GTGGCTCCTAAGCCTGAAGACATTTTTACCAACATAGAGGCAGATCTCAG
AGCAGATCCCTCTTGGTTAATGCAAGACCAAATTTCCTTGCAGCTGAACA
GAACTGCCACCACATTCAGTACATTACAGATCCAGTACTCCAGTGATGCT
CAAATCACTTTACCAAGAGCAGAGATGAGGCCAGTGAAACACAAATGGA
CTATGATTTCAAGGGATAACAATACTAAGCTGGAACATACTGTCTTGGTA
GGTGGAACCGTTGGCCTGAACTGCCCAGGCCAAGGAGACCCCACCCCACA
CGTGGATTGGCTTCTAGCTGATGGAAGTAAAGTGAGAGCCCCTTATGTCA
GTGAGGATGGACGGATCCTAATAGACAAAAGTGGAAAATTGGAACTCCA
GATGGCTGATAGTTTTGACACAGGCGTATATCACTGTATAAGCAGCAATT
ATGATGATGCAGATATTCTCACCTATAGGATAACTGTGGTAGAACCTTTGG
TCGAAGCCTATCAGGAAAATGGGATTCATCACACAGTTTTCATTGGTGAA
ACACTTGATCTTCCATGCCATTCTACTGGTATCCCAGATGCCTCTATTAGC
TGGGTTATTCCAGGAAACAATGTGCTCTATCAGTCATCAAGAGACAAGAA
AGTTCTAAACAATGGCACATTAAGAATATTACAGGTCACCCCGAAAGACC
AAGGTTATTATCGCTGTGTGTGGCAGCCAACCCATCAGGGGGTTGATTTTTTGA
TTTTCCAAGTTTCAGTCAAGATGAAAGGACAAAGGCCCTTGGAGCATGAT
GGAGAAACAGAGGGATCTGGACTTGATGAGTCCAATCCTATTGCTCATCT
TAAGGAGCCACCAGGTGCACAACTCCGTACATCTGCTCTGATGGAGGCTG
AGGTTGGAAAACACACCTCAAGCACAAGTAAGAGGCACAACTATCGGGA
ATTAACACTCCAGCGACGTGGAGATTCAACACATCGACGTTTTAGGGAGA
ATAGGAGGCATTTCCCTCCCTCTGCTAGGAGAATTGACCCACAACATTGG
GCGGCACTGTTGGAGAAAGCTAAAAAGAATGCTATGCCAGACAAGCGAG
AAAATACCACAGTGAGCCCACCCCCAGTGGTCACCCAACTCCCAAACATA
CCTGGTGAAGAAGACGATTCCTCAGGCATGCTCGCTCTACATGAGGAATT

```
TATGGTCCCGGCCACTAAAGCTTTGAACCTTCCAGCAAGGACAGTGACTG
CTGACTCCAGAACAATATCTGATAGTCCTATGACAAACATAAATTATGGC
ACAGAATTCTCTCCTGTTGTGAATTCACAAATACTACCACCTGAAGAACCC
ACAGATTTCAAACTGTCTACTGCTATTAAAACTACAGCCATGTCAAAGAA
TATAAACCCAACCATGTCAAGCCAAATACAAGGCACAACCAATCAACATT
CATCCACTGTCTTTCCACTGCTACTTGGAGCAACTGAATTTCAGGACTCTG
ACCAGATGGGAAGAGGAAGAGAGCATTTCCAAAGTAGACCCCCAATAAC
AGTAAGGACTATGATCAAAGATGTCAATGTCAAAATGCTTAGTAGCACCA
CCAACAAACTATTATTAGAGTCAGTAAATACCACAAATAGTCATCAGACA
TCTGTAAGAGAAGTGAGTGAACCCAGGCACAATCACTTCTATTCTCACAC
TACTCAAATACTTAGCACCTCCACGTTCCCTTCAGATCCACACACAGCTGC
TCATTCTCAGTTTCCGATCCCTAGAAATAGTACAGTTAACATCCCGCTGTT
CAGACGCTTTGGGAGGCAGAGGAAAATTGGCGGAAGGGGGCGGATTATC
AGCCCATATAGAACTCCAGTTCTGCGACGGCATAGATACAGCATTTTCAG
GTCAACAACCAGAGGTTCTTCTGAAAAAAGCACTACTGCATTCTCAGCCA
CAGTGCTCAATGTGACATGTCTGTCCTGTCTTCCCAGGGAGAGGCTCACCA
CTGCCACAGCAGCATTGTCTTTTCCAAGTGCTGCTCCCATCACCTTCCCCA
AAGCTGACATTGCTAGAGTCCCATCAGAAGAGTCTACAACTCTAGTCCAG
AATCCACTATTACTACTTGAGAACAAACCCAGTGTAGAGAAAACAACACC
CACAATAAAATATTTCAGGACTGAAATTTCCCAAGTGACTCCAACTGGTG
CAGTCATGACATATGCTCCAACATCCATACCCATGGAAAAAACTCACAAA
GTAAACGCCAGTTACCCACGTGTGTCTAGCACCAATGAAGCTAAAAGAGA
TTCAGTGATTACATCGTCACTTTCAGGTGCTATCACCAAGCCACCAATGAC
TATTATAGCCATTACAAGGTTTTCAAGAAGGAAAATTCCCTGGCAACAGA
ACTTTGTAAATAACCATAACCCAAAAGGCAGATTAAGGAATCAACATAAA
GTTAGTTTACAAAAAAGCACAGCTGTGATGCTTCCTAAAACATCTCCTGCT
TTACCACAGAGACAAAGTTCCCCTTTCCATTTCACCACACTTTCAACAAGT
GTGATGCAAATTCCATCTAATACCTTGACTACCGCTCACCACACTACGACC
AAAACACACAATCCTGGAAGTCTTCCAACAAAGAAGGAGCTTCCCTTCCC
ACCCCTTAACCCTATGCTTCCTAGTATTATAAGCAAAGACTCAAGTACAA
AAAGCATCATATCAACGCAAACAGCAATACCAGCAACAACTCCTACCTTC
CCTGCATCTGTCATCACTTATGAAACCCAAACAGAGAGATCTAGAGCACA
AACAATACAAAGAGAACAGGAGCCTCAAAAGAAGAACAGGACTGACCCA
AACATCTCTCCAGACCAGAGTTCTGGCTTCACTACACCCACTGCTATGACA
CCTCCTGCTCTGGCATTCACTCATTCCCCACCAGAAAACACAACTGGGATT
TCAAGCACAATCAGTTTTCATTCAAGAACTCTTAATCTGACAGATGTGATT
GAAGAACTAGCCCAAGCAAGTACTCAGACTTTGAAGAGCACAATTGCTTC
TGAAACAACTTTGTCCAGCAAATCACACCAGAGTACCACAACTAGGAAAG
CATCATTAGACACTCCCATACCACCATTCTTGAGCAGCAGTGCTACTCTAA
TGCCAGTTCCCATCTCCCCTCCCTTTACTCAGAGAGCAGTTACTGACACAC
GTGGCGACTCCCATTTCCGGCTTATGACAAATACAGTGGTCAAGCTGCAC
GAATCCTCAAGGCACAATCTCCAAATGCCAAGTTCACAATTGGAACCACT
CACTTCATCTACCTCTAATCTGTTACATTCTACTCCCATGCCAGCACTAAC
AACAGTTAAATCACAGAATTCCAAATTAACTCCATCTCCCTGGGCAGAAT
ACCAATTTTGGCACAAACCATACTCAGACATTGCTGAAAAAGGCAAAAAG
CCAGAAGTAAGCATGTTGGCTACTACAGGCCTGTCCGAGGCCACCACTCT
TGTTTCAGATTGGGATGGACAGAAGAACACAAAGAAGAGTGACTTTGATA
AGAAACCAGTTCAAGAAGCAACAACTTCCAAACTCCTTCCCTTTGACTCTT
TGTCTAGGTATATATTTGAAAAGCCCAGGATAGTTGGAGGAAAAGCTGCA
AGTTTTACTATTCCAGCTAACTCAGATGCCTTTCTTCCCTGTGAAGCTGTT
```

GGAAATCCCCTGCCCACCATTCATTGGACCAGAGTTTCAGGACTTGATTTA
TCTAGAGGAAACCAGAATAGCAGGGTCCAGGTTCTCCCCAATGGTACCCT
GTCCATCCAGAGGGTGGAAATTCAGGACCGCGGACAGTACTTGTGTTCCG
CATCCAATCTGTTTGGCACAGACCACCTTCATGTCACCTTGTCTGTGGTTT
CCTATCCTCCCAGGATCCTGGAGAGACGTACCAAAGAGATCACAGTTCAT
TCCGGAAGCACTGTGGAACTGAAGTGCAGAGCAGAAGGTAGGCCAAGCC
CTACAGTTACCTGGATTCTTGCAAACCAAACAGTTGTCTCAGAATCATCCC
AGGGAAGTAGGCAGGCTGTGGTGACGGTTGACGGAACATTGGTCCTCCAC
AATCTCAGTATTTATGACCGTGGCTTTTACAAATGTGTGGCCAGCAACCCA
GGTGGCCAGGATTCACTGCTGGTTAAAATACAAGTCATTGCAGCACCACC
TGTTATTCTAGAGCAAAGGAGGCAAGTCATTGTAGGCACTTGGGGTGAAA
GTTTAAAACTGCCCTGTACTGCAAAAGGAACTCCTCAGCCCAGCGTTTACT
GGGTCCTCTCTGATGGCACTGAAGTGAAACCATTACAGTTTACCAATTCCA
AGTTGTTCTTATTTTCAAATGGGACTTTGTATATAAGAAACCTAGCCTCTT
CAGACAGGGGCACTTATGAATGCATTGCTACCAGTTCCACTGGTTCGGAG
CGAAGAGTAGTAATGCTTACAATGGAAGAGCGAGTGACCAGCCCCAGGA
TAGAAGCTGCATCCCAGAAAAGGACTGAAGTGAATTTTGGGGACAAATTA
CTACTGAACTGCTCAGCCACTGGGGAGCCCAAACCCCAAATAATGTGGAG
GTTACCATCCAAGGCTGTGGTCGACCAGTGGAGCTGGATCCACGTCTACC
CTAATGGATCCCTGTTTATTGGATCAGTAACAGAAAAAGACAGTGGTGTC
TACTTGTGTGTGGCAAGAAACAAAATGGGGGATGATCTGATACTGATGCA
TGTTAGCCTAAGACTGAAACCTGCCAAAATTGACCACAAGCAGTATTTTA
GAAAGCAAGTGCTCCATGGGAAAGATTTCCAAGTAGATTGCAAAGCTTCC
GGCTCCCCAGTGCCAGAGATATCTTGGAGTTTGCCTGATGGAACCATGAT
CAACAATGCAATGCAAGCCGATGACAGTGGCCACAGGACTAGGAGATAT
ACCCTTTTCAACAATGGAACTTTATACTTCAACAAAGTTGGGGTAGCGGA
GGAAGGAGATTATACTTGCTATGCCCAGAACACCCTAGGGAAAGATGAA
ATGAAGGTCCACTTAACAGTTATAACAGCTGCTCCCCGGATAAGGCAGAG
TAACAAAACCAACAAGAGAATCAAAGCTGGAGACACAGCTGTCCTTGACT
GTGAGGTCACTGGGGATCCCAAACCAAAAATATTTTGGTTGCTGCCTTCC
AATGACATGATTTCCTTCTCCATTGATAGGTACACATTTCATGCCAATGGG
TCTTTGACCATCAACAAAGTGAAACTGCTCGATTCTGGAGAGTACGTATG
TGTAGCCCGAAATCCCAGTGGGGATGACACCAAAATGTACAAACTGGATG
TGGTCTCTAAACCTCCATTAATCAATGGTCTGTATACAAACAGAACTGTTA
TTAAAGCCACAGCTGTGAGACATTCCAAAAAACACTTTGACTGCAGAGCT
GAAGGGACACCATCTCCTGAAGTCATGTGGATCATGCCAGACAATATTTT
CCTCACAGCCCCATACTATGGAAGCAGAATCACAGTCCATAAAAATGGAA
CCTTGGAAATTAGGAATGTGAGGCTTTCAGATTCAGCCGACTTTATCTGTG
TGGCCCGAAATGAAGGTGGAGAGAGCGTGTTGGTAGTACAGTTAGAAGTA
CTGGAAATGCTGAGAAGACCGACATTTAGAAATCCATTTAATGAAAAAAT
AGTTGCCCAGCTGGGAAAGTCCACAGCATTGAATTGCTCTGTTGATGGTA
ACCCACCACCTGAAATAATCTGGATTTTACCAAATGGCACACGATTTTCCA
ATGGACCACAAAGTTATCAGTATCTGATAGCAAGCAATGGTTCTTTTATCA
TTTCTAAAACAACTCGGGAGGATGCAGGAAAATATCGCTGTGCAGCTAGG
AATAAAGTTGGCTATATTGAGAAATTAGTCATATTAGAAATTGGCCAGAA
GCCAGTTATTCTTACCTATGCACCAGGGACAGTAAAAGGCATCAGTGGAG
AATCTCTATCACTGCATTGTGTGTCTGATGGAATCCCTAAGCCAAATATCA
AATGGACTATGCCAAGTGGTTATGTAGTAGACAGGCCTCAAATTAATGGG
AAATACATATTGCATGACAATGGCACCTTAGTCATTAAAGAAGCAACAGC
TTATGACAGAGGAAACTATATCTGTAAGGCTCAAAATAGTGTTGGTCATA

CACTGATTACTGTTCCAGTAATGATTGTAGCCTACCCTCCCCGAATTACAA
ATCGTCCACCCAGGAGTATTGTCACCAGGACAGGGGCAGCCTTTCAGCTC
CACTGTGTGGCCTTGGGAGTTCCCAAGCCAGAAATCACATGGGAGATGCC
TGACCACTCCCTTCTCTCAACGGCAAGTAAAGAGAGGACACATGGAAGTG
AGCAGCTTCACTTACAAGGTACCCTAGTCATTCAGAATCCCCAAACCTCC
GATTCTGGGATATACAAATGCACAGCAAAGAACCCACTTGGTAGTGATTA
TGCAGCAACGTATATTCAAGTAATCTGA

Figure 13

A

Region						
General	Rat	Human	Region Length	% identity	% positives	% gaps
1-655	1-655	1-653	655	76	86	0
656-726	656-726	654-724	71	46	62	0
727-779	727-779	725-777	53	77	86	0
780-1634	780-1617	778-1612	655	38	53	4
1635-end	1618-end	1613-end	980	74	85	0
	Total	Alignment	2614	62	74	1

B

Region					
Rat	Mouse	Length	% identity	% positives	% gaps
1-238	1-238	238	91	92	1

C

Region				
General	Rat	Human	Region Length	% identity
1-1965	1-1965	1-1965	1965	83
1966-2178	1966-2178	1966-2178	213	86
2179-2337	2179-2337	2179-2337	159	86
2338-4893	2338-4893	2338-4863	2565	63
4894-7833	4852-7791	4864-7761	2940	80
	Total	Alignment	7842	80

D

Region				
General	Rat	Mouse	Region Length	% identity
1-720	1-718	1-720	720	93

Figure 14

```
rat_cDNA              CGAGAGACGACAGAAGGTTACGGCTGCGAGAAGACGACAGAAGGGTCCAGAAAAAGGAAA
human_5+3_corrected   ------------------------------------------------------------
mus_cDNA_5            -------------------------------------CAGAAGGGTCCAGGAAA-GGAAA


rat_cDNA              GTGCTGGAGGGGAGTGGGGACAAAAGCAGCGACCAAGTGAATGTCACTTCAGTGACTGAG
human_5+3_corrected   ------------------------------------------------------------
mus_cDNA_5            GTACTGGAGGGGAGTTGGGACAAAAGCAGCGACCAAGGGAACATCGCTTCAGTGACTGAA


rat_cDNA              GCCAGGCAAAACGCGCGGGAAGGATTTTGTGTAGCTTGGGACCCTTTCATAGACACTGAT
human_5+3_corrected   ------------------------------------------------------------
mus_cDNA_5            GCCAGGCAAAAGGAGCGGGAAGGATTATATGTAGCCTGGGACGCTTTCATAAACACTGAT


rat_cDNA              GACACGTTTACGCAAAATA-GAAATTTGAGGAGAAACGCCTGGGCCTTCGGAAAG---GA
human_5+3_corrected   ------------------------------------------------------------
mus_cDNA_5            GACGTGTTTGTGCAAAGCAAGCAATTTGAGGAGAAACGCCTGGGACGTCGGAAAGAAGGA


rat_cDNA              GTGATTGATTAGTACTTGCAAGTTTAGGTGACTTTAAGGAGAACTAACTAATGTATACTA
human_5+3_corrected   ------------------------------------------------------------
mus_cDNA_5            GTGATCGATTAGTACTTGTAAGTTTAGGTGAGTTT---GAGAACTAACTAACCTATACTA


rat_cDNA              TTGAGGGAGGAGGAAGAGCATTACAGAGTTTCCAGCAGCAGCAGGAAAGCTTTGGTTAAT
human_5+3_corrected   ------------------------------------------------------------
mus_cDNA_5            TTGAGGGAGAAGGAAGAGCATT---------CCAGCAGCAGCAGGAAAGCTTTGGTTAGT


rat_cDNA              TTGGAAATGGATGATAGCATTAAAATAACAGAAGCGCCTCCAGGTCTCTGAAGCTTCAGT
human_5+3_corrected   ------------------------------------------------------------
mus_cDNA_5            TTGGAAATGTATGATACCATTAAAATAACAGAAGCGCCTCCAGTTCTCTGAAGAGTCAGT


rat_cDNA              CCCCCAGCTGAAAGCCAGAAAAGACTAAGCCCACTAAGCCTTTTGATCCCTTTGGAAGCA
human_5+3_corrected   ------------------------------------------------------------
mus_cDNA_5            CCCCCAGCT----------A-GTGTAAGCCTACTAAGCCTTTTGCTCCCGTTGGAAGCA


rat_cDNA              AAGAACTTTCCTTCCCTGGGGTGAAGACTCTCCTCAGAAGAITTCCTGTCTCTGCCTATG
human_5+3_corrected   ------------------------------------------------------------
mus_cDNA_5            AAGAACGTTCCTTCAATCAGGTGAAGGCTCTCCTCAGAAGATTTCCTGTCTCTGCTTATG


rat_cDNA              TTACAAGAGGAATCAAAACCAAGACAGAAGAGCTCAGGATGCAGGTGAGAGGCAGGGAAG
human_5+3_corrected   -----------------------------------ATGAAGGTAAAAGGCAGAGGAA
mus_cDNA_5            TTACAAGAGGATTCAAAAGCAAGACAGAAGAGCTCAGGATGCAGAAGAGAGGCAGGGAAG
                                                         ***  **   *  ******  *  *


rat_cDNA              TCAGCGGCTTGTTGATCTCCCTCACTGCTGTCTGCCTGGTGGTCACCCCTGGGAGCAGGG
human_5+3_corrected   TCACCTGCTTGCTGGTCTCCTTTGCTGTGATCTGCCTGGTCGCCACCCCTGGGGGCAAGG
mus_cDNA_5            TCAGCTGCTTGCTGATCTCCCTCACTGCCATCTGCCTGGTGGTCACCCCTGGGAGCAGGG
                      ***  *  ***** **  ***** *   ***  *********  * *********  *** **


rat_cDNA              CCTGTCCTCGCCGCTGTGCCTGCTATGTGCCCACAGAGGTGCACTGTACATTCGGTACC
human_5+3_corrected   CCTGTCCTCGCCGCTGTGCCTGTTATATGCCTACGGAGGTGCACTGCACATTCGGTACC
mus_cDNA_5            TCTGTCCTCGCCGATGTGCCTGCTATGTGCCCACAGAGGTGCACTGTACATTCGGGACC
                      ************** *******  ***  **** **  ***** ****** *********  ***


rat_cDNA              TGACCTCCATCCCAGATGG-CATCCCGGCCAATGTGGAACGAATAAATTTAGGATATAAC
human_5+3_corrected   TGACTTCCATCCCAGACAG-CATCCCGCCCAATGTGGAACGCATCAATTTAGGATACAAC
mus_cDNA_5            TGACCTCCATCCCAGACGGGCATCCCAGCCAATGTGGAACGAGTCAATTTAGGGTATAAC
                      ****  *********** **  *  ******  *************  **  ********* ** ***
```

```
rat_cDNA              AGCCTTACTAGATTGACAGAAAACGACTTTGATGGCCTGAGCAAACTGGAGTTACTCATG
human_5+3_corrected   AGCTTGGTTAGATTGATGGAAACAGATTTTTCTGGCCTGACCAAACTGGAGTTACTCATG
mus_cDNA_5            AGCCTCACTAGATTGACAGAAAATGACTTTTCTGGCCTGAGCAGACTGGAGTTACTCATG
                     *** * ******** **** ** *** ******* ** ****************


rat_cDNA             .CTGCACAGTAATGGCATTCACAGAGTCAGTGACAAGACCTTCTCGGGCTTGCAGTCCTTG
human_5+3_corrected  CTTCACAGCAATGGCATTCACACAATCCCTGACAAGACCTTCTCAGATTTGCAGGCCTTG
mus_cDNA_5           CTGCACAGCAATGGCATTCACAGAGTCAGTGACAAGACCTTCTCGGGCTTGCAGTCCTTG
                     ** ***** ************* * ** ************** * ****** *****


rat_cDNA             CAGGTCTTAAAAATGAGCTATAACAAAGTCCAAATCATTCGGAAGGATACTTTCTACGGA
human_5+3_corrected  CAGGTCTTAAAAATGAGCTATAATAAAGTCCGAAAACTTCAGAAAGATACTTTTTATGGC
mus_cDNA_5           CAGGTCTTAAAAATGAGCTATAACAAAGTCCAAATAATTGAGAAGGATACTTTGTATGGA
                     ********************** ** ****** ** ** *** ******* ** **


rat_cDNA             CTCGGGAGCTTGGTCCGGTTGCACCTGGATCACAACAACATTGAATTCATCAACCCTGAG
human_5+3_corrected  CTCAGGAGCTTGACACGATTGCACATGGACCACAACAATATTGAGTTTATAAACCCAGAG
mus_cDNA_5           CTCAGGAGCTTGACCCGGTTGCACCTGGATCACAACAACATTGAGTTTATCAACCCCGAG
                     *** ******** ** ****** **** ******** ***** ** ** ***** ***


rat_cDNA             GCCTTTTATGGACTTACCTCGCTCCGCTTGGTACATTTAGAAGGAAACCGGCTCACAAAG
human_5+3_corrected  GTTTTTTATGGGCTCAACTTTCTCCGCCTGGTGCACTTGGAAGGAAATCAGCTCACTAAG
mus_cDNA_5           GCGTTTTACGGACTCACCTTGCTCCGCTTGGTACATCTAGAAGGAAACCGGCTGACAAAG
                     *  ***** ** ** ** ** * ** ****** ** * ******** * *** ** ***


rat_cDNA             CTCCATCCAGACACATTTGTCTCATTAAGCTATCTCCAGATATTTAAAAACCTCTTTCATT
human_5+3_corrected  CTCCACCCAGATACATTTGTCTCTTTGAGCTACCTCCAGATATTTAAAAATCTCTTTCATT
mus_cDNA_5           CTCCATCCAGACACATTTGTCTCTTTGAGCTATCTCCAGATATTTAAAAACCTCCTTCATT
                     ***** ***** *********** ** ***** ****************** *** ******


rat_cDNA             AAGTACCTGTTCTTGTCTGATAACTTCCT-GACCTCCCTCCCAAAAGAAATGGTCTCCTA
human_5+3_corrected  AAGTTCCTATACTTGTCTGATAACTTCCT-GACCTCCCTCCCTCAAGAGATGGTCTCCTA
mus_cDNA_5           AAGNACCTGTACTTGTATGATAACTTCATTGACCTCCCTCCCAAAAGAAATGGTCTCCTC
                     *** ***  * ****** ******** * * ***********  **** **********


rat_cDNA             CATGCCAAACCTAGAAAGCCTGTATTTGCATGGAAACCCATGGACCTGTGACTGCCATTT
human_5+3_corrected  TATGCCTGACCTAGACAGCCTTTACCTGCATGGAAACCCATGGACCTGTGATTGCCATTT
mus_cDNA_5           TATGCCAAACCTAGAAAGCCTTTACTTGCATGGAAACCCATGGACCTGTGACTGCCATTT
                     *****  ****** ***** ** ** *** ******************* ********


rat_cDNA             AAAGTGGTTGTCTGAGTGGATGCAGGGAAACCCAGATATAATAAAAATGCAAGAAAGACAG
human_5+3_corrected  AAAGTGGTTGTCTGACTGGATACAGGGNNNNNCCAGATGTAATAAAAATGCAAAAAAGATAG
mus_cDNA_5           AAAGTGGTTGTCCGAGTGGATGCAGGGAAACCCAGGTA-ACTATCTTGT-----TTGTTTG
                     ************ ** ***** **** **** * * ** ** * *


rat_cDNA             AAGCTCTTCCAGTCCTCAGCAATGTCCCCT-TTGCATGAACCCCAGGATCTCTAAAGGCA
human_5+3_corrected  AAGTCCCTCTAGTGCTCAGCAGTGTCCACT-TTGCATGAACCCTAGGACTTCTAAAGGCA
mus_cDNA_5           TTTCTTTTTTTATARKACGTATTTTCCTCAATTTCATTTAGAATGATATCCCAAAAGTC-
                           *          *    * * * *** *  ** *** *        *     * **** *


rat_cDNA             GACCCTTTGCTATGGTACCATCTGGAGCTTTCCTATGTACAAAGCCAACCATTGATCCAT
human_5+3_corrected  AGCCGTTAGCTATGGTCTCAGCTGCAGCTTTCCAGTGTGCCAAGCCAACCATTGACTCAT
mus_cDNA_5           -CCCCATAACCTCCCCCCCA--------CTTCCCTACCTACCCATTC---CCATTTTTTGGC
                      ** * *   **          **    *** * * *   * **** *


rat_cDNA             CACTGAAGTCAAAGAGCCTGGTTACTCAGGAGGACAATGGATCTGCCTCCACCTCACCTC
human_5+3_corrected  CCCTGAAATCAAAGAGCCTGACTATTCTGGAAGACAGTAGTTCTGCTTTCATCTCTCCCC
mus_cDNA_5           CCTGGCATTCCCC-----------------------------------------------
                     *   * * **


rat_cDNA             AAGATTTCATAGAACCCTTTGGCTCCTTGTCTTTGAACATGACANANNTNTCTGGAAATA
human_5+3_corrected  AAGGTTTCATGGCACCCTTTGGCTCCCTCACTTTGAATATGACAGATCAGTCTGGAAATG
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA             AGGCCGACATGGTCTGTAGTATCCAAAAGCCATCAAGGACATCACCAACTGCATTCACTG
human_5+3_corrected  AAGCTAACATGGTCTGCAGTATTCAAAAGCCCTCAAGGACATCACCCATTGCATTCACTG
mus_cDNA_5           ------------------------------------------------------------
```

```
rat_cDNA              AAGAAAATGACTACATCATGCTAAATGCGTCATTTTCCACAAATCTTGTGTGCAGTGTAG
human_5+3_corrected   AAGAAAATGACTACATCGTGCTAAATACTTCATTTTCAACATTTTTGGTGTGCAACATAG
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              ATTATAATCACATCCAGCCAGTGTGGCAACTTCTGGCTTTATACAGTGACTCTCCTCTGA
human_5+3_corrected   ATTACGGTCACATTCAGCCAGTGTGGCAAATTTTGGCTTTGTACAGTGATTCTCCTCTGA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              TACTAGAAAGGAAGCCCCAGCTTACCGAGACTCCTTCACTGTCTTCTAGATATAAACAGG
human_5+3_corrected   TACTAGAAAGGAGCCACTTGCTTAGTGAAACACCGCAGCTCTATTACAAATATAAACAGG
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              TGGCTCTTAGGCCTGAAGACATTTTTACCAGCATAGAGGCTGATGTCAGAGCAGACCCTT
human_5+3_corrected   TGGCTCCTAAGCCTGAAGACATTTTTACCAACATAGAGGCAGATCTCAGAGCAGATCCCT
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              TTTGGTTCCAACAAGAAAAAATTGTCTTGCAGCTGAACAGAACTGCCACCACACTTAGCA
human_5+3_corrected   CTTGGTTAATGCAAGACCAAATTTCCTTGCAGCTGAACAGAACTGCCACCACATTCAGTA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              CATTACAGATCCAGTTTTCCACTGATGCTCAAATCGCTTTACCAAGGGCGGAGATGAGAG
human_5+3_corrected   CATTACAGATCCAGTACTCCAGTGATGCTCAAATCACTTTACCAAGAGCAGAGATGAGGC
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              CGGAGAGACTCAAATGGACCATGATCCTGATGATGAACAATCCCAAACTGGAACGCACTG
human_5+3_corrected   CAGTGAAACACAAATGGACTATGATTTCAAGGGATAACAATACTAAGCTGGAACATACTG
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              TCCTGGTTGGCGGCACTATTGCCCTGAGCTGTCCAGGCAAAGGCGACCCTTCACCTCACT
human_5+3_corrected   TCTTGGTAGGTGGAACCGTTGGCCTGAACTGCCCAGGCCAAGGAGACCCCACCCCACACG
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              TGGAATGGCTTCTAGCTGATGGGAGTAAAGTGAGAGCCCCTTACGTTAGCGAGGATGGGC
human_5+3_corrected   TGGATTGGCTTCTAGCTGATGGAAGTAAAGTGAGAGCCCCCTTATGTCAGTGAGGATGGAC
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              GAATCCTAATAGACAAAAATGGGAAGTTGGAACTGCAGATGGCTGACAGCTTTGATGCAG
human_5+3_corrected   GGATCCTAATAGACAAAAGTGGAAAATTGGAACTCCAGATGGCTGATAGTTTTGACACAG
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              GTCTTTACCACTGCATAAGCACCAATGATGCAGATGCGGATGTTCTCACATACAGGATAA
human_5+3_corrected   GCGTATATCACTGTATAAGCAGCAATTATGATGATGCAGATATTCTCACCTATAGGATAA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              CTGTGGTAGAGCCCTATGGAGAAAGCACACATGACAGTGGAGTCCAGCACACAGTGGTTA
human_5+3_corrected   CTGTGGTAGAACCTTTGGTCGAAGCCTATCAGGAAAATGGGATTCATCACACAGTTTTCA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              CGGGTGAGACGCTCGACCTTCCATGCCTTTCCACGGGTGTTCCAGATGCTTCTATTAGCT
human_5+3_corrected   TTGGTGAAACACTTGATCTTCCATGCCATTCTACTGGTATCCCAGATGCCTCTATTAGCT
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              GGATTCTTCCAGGGAACACTGTGTTCTCTCAGCCATCAAGAGACAGGCAAATTCTTAACA
human_5+3_corrected   GGGTTATTCCAGGAAACAATGTGCTCTATCAGTCATCAAGAGACAAGAAAGTTCTAAACA
mus_cDNA_5            ------------------------------------------------------------
```

```
rat_cDNA                ATGGGACCTTAAGAATATTACAGGTTACGCCAAAAGATCAAGGTCATTACCAATGTGTGG
human_5+3_corrected     ATGGCACATTAAGAATATTACAGGTCACCCCGAAAGACCAAGGTTATTATCGCTGTGTGG
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                CTGCCAACCCATCAGGGGCCGACTTTTCCAGTTTTAAAGTTTCAGTTCAAAAGAAAGGCC
human_5+3_corrected     CAGCCAACCCATCAGGGGTTGATTTTTTGATTTTCCAAGTTTCAGTCAAGATGAAAGGAC
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA                AAAGGATGGTTGAGCATGACAGGGAGGCAGGTGGATCTGGACTTGGAGAACCCAACTCCA
human_5+3_corrected     AAAGGCCCTTGGAGCATGATGGAGAAACAGAGGGATCTGGACTTGATGAGTCCAATCCTA
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA                GTGTTTCCCTTAAGCAGCCAGCATCTTTGAAACTCTCTGCATCAGCTTTGACAGGGTCAG
human_5+3_corrected     TTGCTCATCTTAAGGAGCCACCAGGTGCACAACTCCGTACATCTGCTCTGATGGAGGCTG
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA                AGGCTGGAAAACAAGTCTCCGGTGTACATAGGAAGAACAAACATAGAGACTTAATACATC
human_5+3_corrected     AGGTTGGAAAACACACCTCAAGCACAAGTAAGAGGCACAACTATCGGGAATTAACACTCC
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA                GGCGGCGTGGGGATTCCACGCTCCGGCGATTCAGGGAGCATAGGAGGCAGCTCCCTCTCT
human_5+3_corrected     AGCGACGTGGAGATTCAACACATCGACGTTTTAGGGAGAATAGGAGGCATTTCCCTCCCT
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA                CTGCTCGGAGAATTGACCCGCAACGCTGGGCAGCACTTCTAGAAAAAGCCAAAAAGAATT
human_5+3_corrected     CTGCTAGGAGAATTGACCCACAACATTGGGCGGCACTGTTGGAGAAAGCTAAAAAGAATG
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA                CTGTGCCAAAAAAGCAAGAAAATACCACAGTAAAGCCAGTGCCACTGGCTGTTCCCCTCG
human_5+3_corrected     CTATGCCAGACAAGCGAGAAAATACCACAGTGAGCCCACCCCCAGTGGTCACCCAACTCC
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA                TGGAACTCACTGACGAGGAAAAGGATGCCTCTGGCATGATTCCTCCAGATGAAGAATTCA
human_5+3_corrected     CAAACATACCTGGTGAAGAAGACGATTCCTCAGGCATGCTCGCTCTACATGAGGAATTTA
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA                TGGTTCTGAAAACTAAGGCTTCTGGTGTCCCAGGAAGGTCACCAACTGCTGACTCTGGAC
human_5+3_corrected     TGGTCCCGGCCACTAAAGCTTTGAACCTTCCAGCAAGGACAGTGACTGCTGACTCCAGAA
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA                CAGTAAATCATGGTTTTATGACGAGTATAGCTTCTGGCACAGAAGTCTCAACTGTGAATC
human_5+3_corrected     CAATATCTGATAGTCCTATGACAAACATAAATTATGGCACAGAACTCTCCGTTGTGAATT
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA                CACAAACACTACAATCTGAGCACCTTCCTGATTTCAAATTATTTAGTGTAACAAACGGTA
human_5+3_corrected     CACAAATACTACCACCTGAAGAACCCACAGATTTCAAACTGTCTACTGCTATTAAAACTA
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA                CAGCTGTGACAAAGAGTATGAACCCATCCATAGCAAGCAAAATAGAAGATACAACCAACC
human_5+3_corrected     CAGCCATGTCAAAGAATATAAACCCAACCATGTCAAGCCAAATACAAGGCACAACCAATC
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA                AAAACCCAATCATTATCTTTCCATC---------AGTAGCTGAAATTCGAGATTCTG-CT
human_5+3_corrected     AACATTCATCCACTGTCTTTCCACTGCTACTTGGAGCAACTGAATTTCAGGACTCTGACA
mus_cDNA_5             ------------------------------------------------------------
```

```
rat_cDNA              ·CAGGCA----GGAAGAGCATCTTCCCAAAGTGCACACCCTGTAACAGGGGGAAACATGGCT
human_5+3_corrected   GAGGGAAGAGGAAGAGAGCATTTCC--AGTA-ACCCCCAATAACAGTAAGGACTATGATC
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              ACCTATGGCCATACCAACACATA--TAGTAGCTTTACCAGCAAAGCCAGTACAGTCTTGC
human_5+3_corrected   AAAGATGNTCAATGTCAAANATGCTTAGTAGCACCACCAACAAA-CTATTA------TTAG
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              AGCCAATAAATCCAACAGAAAGTTATGGACCTCAGATACCTATTACAGGAGTCAGCAGAC
human_5+3_corrected   AGTCAGTAAATACCACAAATAGTCAT------CAGACATCTGTAAGAGAAGTGAGTGAAC
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              CTAGCAGTAGTGACATCTCTTCTCACACTACTGCAGACCCTAGCTTCTCCAGTCACCCTT
human_5+3_corrected   CCAGGCACAATCACTTCTATTCTCACACTACTCAAATACTTAGCACCTCCACGTTCCCTT
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              CAGGTTCACACACCACTGCCTCGTCTTTATTTCACATTCCTAGAAACAACAATACAGGTA
human_5+3_corrected   CAGATCCACACACACAGCTGCTCATTCTCAGTTTCCGATCCCTAGANNNAATAGTACAGTTA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              ACTTCCCCTTGTCCAGGCACTTGGGAAGAGAGAGGACAATTTGGAGCAGAGGGAGAGTTA
human_5+3_corrected   ACATCCCGCTGTTCAGACGCTTTGGGAGGCAGAGGAAAATTGGCGGAAGGGGGCGGATTA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              AAAACCCACATAGAACCCCAGTTCTCCGACGGCATAGACACAGGACTGTGAGGCCAGCAA
human_5+3_corrected   TCAGCCCATATAGAACTCCAGTTCTGCGACGGCATAGATACAGCATTTTCAGGTCAACAA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              TCAAGGGACCTGCTAACAAAAATGTGAGCCAAGTTCCAGCCACAGAGTACCCTGGGATGT
human_5+3_corrected   CCAGAGGTTCTTCTGAAAAAAGCACTACTGCATTCTCAGCCACAGTGCTCAATGTGACAT
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              GCCACACATGTCCTTCCGCAGAGGGGCTCACAGTGGCTACTGCAGCACTGTCAGTTCCAA
human_5+3_corrected   GTCTGTCCTGTCTTCCCAGGGAGAGGCTCACCACTGCCACAGCAGCATTGTCTTTTCCAA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              GTTCATCCCACAGTGCCCTCCCCAAAACTAATAATGTTGGGGTCATAGCAGAAGAGTCTA
human_5+3_corrected   GTGCTGCTCCCATCACCTTCCCCAAAGCTGACATTGCTAGAGTCCCATCAGAAGAGTCTA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              CCACTGTGGTCAAGAAACCACTGTTACTATTTAAGGACAAACAAAATGTAGATATTGAGA
human_5+3_corrected   CAACTCTAGTCCAGAATCCACTATTACTACTTGAGAACAAACCCAGTGTAGANNNNGAAA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              TAATAACAACCACTACAAAAATATTCCGGAGGGGAAAGTAACCACGTGATTCCTACGGAAG
human_5+3_corrected   NNACAACACCCACAATAAAAATATTCAGGACTNGAAATTTCCCAAGTGACTCCAACTGGTG
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              CAAGCATGACTTCTGCTCCAACATCTGTATCCCTGGGGAAATCTCCTGTAGACAATAGTG
human_5+3_corrected   CAGTCATGACATATGCTCCAACATCCATACCCATGGAAAAAACTCACAAAGTAAACGCCA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              GTCACCTGAGCATGCCTGGGACCATCCAAACTGGGAAAGATTCAGTGGAAACAACACCAC
human_5+3_corrected   GTTACCCACGTGTGTCTAGCACCAATGAAGCTAAAAGAGATTCAGTGATTACATCGTCAC
mus_cDNA_5            ------------------------------------------------------------
```

```
rat_cDNA            TTCCCAGCCCCCTCAGCACACCCTCAATACCAA------CAAGCACAAAATTCTCAAAGA
human_5+3_corrected TTTCAGGTGCTATCACCAAGCCACCAATGACTATTATAGCCATTACAAGGTTTTCAAGAA
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            GGAAAACTCCCTTGCACCAGATCTTTGTAAATAACCAGAAGAAGGAGGGGATGTTAAAGA
human_5+3_corrected GGAAAATTCCCTGGCAACAGAACTTTGTAAATAACCATAACCCAAAAGGCAGATTAAGGA
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            ATCCATATCAATTCGGTTTACAAAAGAACCCAGCCGCAAAGCTTCCCAAAATAGCTCCTC
human_5+3_corrected ATCAACATAAAGTTAGTTTACAAAAAAGCACAGCTGTGATGCTTCCTAAAACATCTCCTG
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            TTTTACCCACAG-GTCAGAGTTCCCCCTCAGATTCTACAACTCTCTTGACAAGTCCGCCA
human_5+3_corrected CTTTACC-ACAGAGACAAAGTTCCCCTTTCCATTCACCACACTTTCAACAAGTGTGATG
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            CCAGCTCTGTCTACAACAATGGCTGCCACTCAGAACAAGGGCACTGAAGTAGTATCAGGT
human_5+3_corrected CAAATTCCATCTAATACCTTGACTACCGGCTCACCACACTACGACCAAA--ACACACAA-T
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            GCCAGAAGTCTCTCAGCAGGGAAGAAGCAGCCCTTCACCAACTCCTCTCCAG-TGCTTCC
human_5+3_corrected CCTGGAAGTCTTCCAACAAAGAAGGAGCTTCCCTTC-CCACCCCTTAACCCTATGCTTCC
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            TAGCACCATAAGCAAGAGATCTAATACATTAAACTTCTTGTCAACGGAAACCCCCACAGT
human_5+3_corrected TAGTATTATAAGCAAAGACTCAAGTACAAAAAGCATCATATCAACGCAAACAGCAACCGC
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            GACAAGTCCTACTGCTACTGCATCTGTCATTATGTCTGAAACCCAACGAACAAGATCCAA
human_5+3_corrected AACAACTCCTACCTTCCCTGCATCTGTCATCACTTATGAAACCCAAACAGAGAGATCTAG
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            AGAAGCAAAAGACCAAATAAAGGGGCCTCG---GAAGAACAGAAACAACGCAAACACCAC
human_5+3_corrected AGCACAAACAATACAAAGAGAAGGACCTCAAAAGAAGAACAGGACTGACCCAAACATCTC
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            CCCCAGGCAGGTTTCTGGCTATAGTGCATACTCAGCTCTAACAACAGCTGATACCCCCTT
human_5+3_corrected TCCAGACCAGAGTTCTGGCTTCACTACACCCACTGCTATGACNAGCTCCTCCTNNGCTCTNNN
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            GGCTTTCAGTCATTCCCCACGACAAGATGATGGTGGAAATGTAAGTGCAGTTGCTTATCA
human_5+3_corrected NGCATTCACTCATTCCCCACCAGAAAAACACAACTGGGATTTCAAGCACAATCAGTTTTCA
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            CTCAACAACCTCTCTTCTGGCCA---TAACTGAACTGTTTGAG-AAGTAC--ACCCAGAC
human_5+3_corrected TTCAAGAACTCTTAATCTGACAGATGTGATTGAAGAACTAGCCCAAGCAAGTACTCAGAC
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            TTTGGGAAATACAACAGCTTTGGAAACAACGTTGTTGAGCAAATCACAGGAGAGTACCAC
human_5+3_corrected TTTGAAGAGCACAATTGCTTCTGAAACAACTTTGTCCAGCAAATCACACCAGAGTACCAC
mus_cDNA_5          ------------------------------------------------------------

rat_cDNA            AGTGAAAAGAGCCTCA---GACAC----ACCACCACCACTCCTCAGCAGTGGGGCGCCCC
human_5+3_corrected AACTAGGAAAGCATCATTAGACACTCAACCACCACCATTCTTGAGCAGCAGTGCTACTCT
mus_cDNA_5          ------------------------------------------------------------
```

```
rat_cDNA               AGTGCCCACTCCTTCCCCACCTCCTTTTACTAAGGGTGTGGTTACAGACAGCAAAGTCAC
human_5+3_corrected    AATGCCAGTTCCCATCTCCCCTCCCTTTACTCAGAGAGCAGTTACTGACAACGTGGCGAC
mus_cDNA_5             ------------------------------------------------------------


rat_cDNA               ATCAGCTTTCCAGATGACGTCAAATAGAGTGGTCACCATATATGAATCTTCAAGGCACAA
human_5+3_corrected    TCCCATTTCCGGGGCTTATGACAAATACAGTGGTCAAGCTGCACGAATCCTCAAGGCACAA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA               TACAGATCTGCAGCAACCCTCAGCAGAGGCTAGCCCCAATCCTGAGATCATAACTGGAAC
human_5+3_corrected    TCCNNNNNNNNNCA--AATGCCAAGTTCA--------CNNAATTGNGAACCNNNNACTCNNNN
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA               CACTGACTCTCCCTCTAATCTGTTTCCATCCACTTCTGTGCCAGCACTAAGGGTAGATAA
human_5+3_corrected    NACTTCATCTACNTCTAATCTGTTACATTCTACTCCCATGCCAGCACTAACAACAGTTAA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA               ACCACAGAATTCTAAATGGAAGCCCTCTCCCTGGCCAGAACACAAATATCAGCTCAAGTC
human_5+3_corrected    ATCACAGAATTCCAAATTAACTCCATCTCCCTGGGCAGAATACCAATTTTGGCACAAACC
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA               ATACTCCGAAACCATTGAGAAGGGCAAAAGGCCAGCAGTAAGCATGTCCCCCCACCTCAG
human_5+3_corrected    ATACTCAGACATTGCTGAAAAAGGCAAAAAGCCAGAAGTAAGCATGTTGGCT-ACTACAG
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA               -CCTTCCAGAGGCCAGCACTCATGCCTCACACTGGAATACACAGAAGCATGCAGAAAAGA
human_5+3_corrected    GCCTGTCCGAGGCCACCACTCTTGTTTCAGATTGGGATGGACAGAAGAACACAAAGAAGA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA               GTGTTTTTGATAAGAAACCTGGTCAA-AACC--CAACTTCCAAACATCTGCCTTACGTCT
human_5+3_corrected    GTGACTTTGATAAGAAACCAGTTCAAGAAGCAACAACTTCCAAACTCCTTCCCTTTGACT
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA               CTCTACCTAAGACTCTATTGAAAAAGCCAAGAATAATTGGAGGAAAAGGCTGCAAGCTTTA
human_5+3_corrected    CTTTGTCTAGGTATATATTTGAAAAGCCCAGGATAGTTGGAGGAAAAGCTGCAAGTTTTA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA               CAGTTCCAGCTAATTCAGACGTTTTTTCTTCCTTGTGAGGCTGTTGGAGACCCACTGCCCA
human_5+3_corrected    CTATTCCAGCTAACTCAGATGCCTTTCTTCCCTGTGAAGCTGTTGGAAATCCCCTGCCCA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA               TCATCCACTGGACCAGAGTTTCATCAGGANTTGAAATATCCCAAGGGACACAGAAAAGCC
human_5+3_corrected    CCATTCATTGGACCAGAGTNNNNNTCAGGACTTGATTTATCTAAGAGGAAACAGAATAGCA
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA               GGTTCCACGTGCTTCCCAATGGCACCTTGTCCATCCAGAGGGTCAGTATTCAGGACCGTG
human_5+3_corrected    GGGTCCAGGTTCTCCCCAATGGTACCCTGTCCATCCAGAGGGTGGAAATTCAGGACCGCG
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA               GACAGTACCTGTGCTCTGCATTTAATCCACTGGGCGTAGACCCATTTTCATGTCTCTTTGT
human_5+3_corrected    GACAGTACTTGTGTTCCGCATCCAATCTGTTTGGCACAGACCCACCTTCATGTCACCTTGT
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA               CTGTGGTTTTTTACCCGGCAAGGATTTTGGACAGACATGTCAAGGAGATCACAGTTCACT
human_5+3_corrected    CTGTGGTTTCCTATCCTCCCAGGATCCTGGAGAGACGTACCAAAGAGATCACAGTTCATT
mus_cDNA_5            ------------------------------------------------------------
```

```
rat_cDNA              TTGGAAGTACTGTGGAACTAAAGTGCAGAGTGGAGGGTATGCCGAGGCCTACGGTTTCCT
human_5+3_corrected   CCGGAAGCACTGTGGAACTGAAGTGCAGAGCAGAAGGTAGGCCAAGCCCTACAGTTACCT
mus_cDNA_5            ------------------------------------------------------------

rat_cDNA              GGATACTTGCAAACCAAACGGTGGTCTCAGAAACGGCCAAGGGAAGCAGAAAGGTCTGGG
human_5+3_corrected   GGATTCTTGCAAACCAAACAGTTGTCTCAGAATCATCCCAGGGAAGTAGGCAGGCTGTGG
mus_cDNA_5           ------------------------------------------------------------

rat_cDNA              TAACACCTGATGGAACATTGATCATCTATAATCTGAGTCTTTATGATCGTGGTTTTTACA
human_5+3_corrected   TGACGGTTGACGGAACATTGGTCCTCCACAATCTCAGTATTTATGACCGTGGCTTTTACA
mus_cDNA_5           ------------------------------------------------------------

rat_cDNA              AGTGTGTGGCCAGCAACCCATCTGGCCAGGATTCACTGTTGGTTAAGATACAAGTCATCA
human_5+3_corrected   AATGTGTGGCCAGCAACCCAGGTGGCCAGGATTCACTGCTGGTTAAAATACAAGTCATTG
mus_cDNA_5           ------------------------------------------------------------

rat_cDNA              CAGCTCCCCTGTCATTATAGAGCAAAAGAGGCAAGCCATCGTTGGGGTTTTAGGTGGAA
human_5+3_corrected   CAGCACCACCTGTTATTCTAGAGCAAAGGAGGCAAGTCATTGTAGGCACTTGGGGTGAAA
mus_cDNA_5.          ------------------------------------------------------------

rat_cDNA              GTTTGAAACTGCCCTGCACTGCAAAAGGAACTCCCCAGCCTAGTGTTCACTGGGTCCTTT
human_5+3_corrected   GTTTAAAACTGCCCTGTACTGCAAAAGGAACTCCTCAGCCCAGCGTTTACTGGGTCCTCT
mus_cDNA_5           ------------------------------------------------------------

rat_cDNA              ATGATGGGACTGAACTAAAACCATTGCAGTTGACTCATTCCAGATTTTTCTTGTATCCAA
human_5+3_corrected   CTGATGGCACTGAAGTGAAACCATTACAGTTTACCAATTCCAAGTTGTTCTTATTTTCAA
mus_cDNA_5           ------------------------------------------------------------

rat_cDNA              ATGGAACTCTGTATATAAGAAGCATCGCTCCTTCAGTGAGGGGCACTTATGAGTGCATTG
human_5+3_corrected   ATGGGACTTTGTATATAAGAAACCTAGCCTCTTCAGACAGGGGCACTTATGAATGCATTG
mus_cDNA_5           ------------------------------------------------------------

rat_cDNA              CCACCAGCTCCTCAGGCTCAGAGAGAAGGGTAGTGATTCTTACTGTGGAAGAGGGAGAGA
human_5+3_corrected   CTACCAGTTCCACTGGTTCGGAGCGAAGAGTAGTAATGCTTACAATGGAAGAGCGAGTGA
mus_cDNA_5           ------------------------------------------------------------

rat_cDNA              CAATCCCCAGGATAGAAACTGCCTCTCAGAAATGGACTGAGGTGAATTTGGGTGAGAAAT
human_5+3_corrected   CCAGCCCCAGGATAGAAGCTGCATCCCAGAAAAGGACTGAAGTGAATTTTGGGGACAAAT
mus_cDNA_5           ------------------------------------------------------------

rat_cDNA              TACTACTGAACTGCTCAGCTACTGGGGATCCAAAGCCTAGAATAATCTGGAGGCTGCCAT
human_5+3_corrected   TACTACTGAACTGCTCCAGCCACTGGGGAGCCCAAACCCCAAATAATGTGGAGGTTACCAT
mus_cDNA_5           ------------------------------------------------------------

rat_cDNA              CCAAGGCTGTCATCGACCAGTGGCACAGAATGGGCAGCCGAATCCACGTCTACCCAAATG
human_5+3_corrected   CCAAGGCTGTGGTCGACCAGTGG----------GCAGCTGGATCCACGTCTACCCTAATG
mus_cDNA_5.          ------------------------------------------------------------

rat_cDNA              GATCCTTGGTGGTTGGGTCAGTGACGGAAAAAGACGCTGGTGACTACTTATGTGTGGCAA
human_5+3_corrected   GATCCCTGTTTATTGGATCAGTAACAGAAAAAGACAGTGGTGTCTACTTGTGTGTGGCAA
mus_cDNA_5           ------------------------------------------------------------

rat_cDNA              GAAACAAAATGGGAGATGACCTAGTCCTGATGCATGTCCGCCTGAGATTGACACCTGCCA
human_5+3_corrected   GAAACAAAATGGGGGATGATCTGATACTGATGCATGTTAGCCTAAGACTGAAACCTGCCA
mus_cDNA_5           ------------------------------------------------------------
```

```
rat_cDNA              AAATTGAACAGAAGCAGTATTTTAAGAAGCAAGTGCTCCATGGGAAAGATTTCCAAGTTG
human_5+3_corrected   AAATTGACCACAAGCAGTATTTTAGAAAGCAAGTGCTCCATGGGAAAGATTTCCAAGTAG
mus_cDNA_5            ------------------------------------------------------------


rat_cDNA              ACTGCAAGGCCTCTGGCTCCCCTGTGCCTGAGGTATCCTGGAGTTTGCCTGATGGGACAG
human_5+3_corrected   ATTGCAAAGCTTCCGGCTCCCCAGTGCCAGAGATATCTTGGAGTTTGCCTGATGGAACCA
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA              TGCTCAACAATGTAGCCCAAGCTGATGACAGTGGCTATAGGACCAAGAGGTACACCCTTT
human_5+3_corrected   TGATCAACAATGCAATGCAAGCCGATGACAGTGGCCACAGGACTAGGAGATATACCCTTT
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA              TCCACAATGGAACCTTGTATTTCAACAACGTTGGGATGGCAGAGGAAGGAGATTATATCT
human_5+3_corrected   TCAACAATGGAACTTTATACTTCAACAAAGTTGGGGTAGCGGAGGAAGGAGATTATACTT
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA              GCTCTGCCCAGAACACCTTAGGGAAAGATGAGATGAATGTCCACCTAACAGTTCTAACAG
human_5+3_corrected   GCTATGCCCAGAACACCCTAGGGAAAGATGAAATGAAGGTCCACTTAACAGTTATAACAG
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA              CCATCCCACGGATAAGGCAAAGCTACAAGACCACCATGAGGCTCAGGGCTGGAGAAACAG
human_5+3_corrected   CTGCTCCCCGGATAAGGCAGAGTAACAAAACCAACAAGAGAATCAAAGCTGGAGACACAG
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA              CTGTCCTTGACTGCGAGGTCACTGGGGAACCGAAGCCCAATGTATTTTGGTTGCTGCCTT
human_5+3_corrected   CTGTCCTTGACTGTGAGGTCACTGGGGATCCCAAACCAAAAATATTTTGGTTGCTGCCTT
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA              CCAACAATGTCATTTCATTCTCCAATGACAGGTTCACATTTCATGCCAATAGAACTTTGT
human_5+3_corrected   CCAATGACATGATTTCCTTCTCCATTGATAGGTACACATTTCATGCCAATGGGTCTTTGA
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA              CCATCCATAAAGTGAAACCACTTGACTCTGGGGACTATGTGTGCGTAGCTCAGAATCCTA
human_5+3_corrected   CCATCAACAAAGTGAAACTGCTCGATTCTGGAGAGTACGTATGTGTAGCCCGAAATCCCA
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA              GTGGGGATGACACTAAGACATACAAACTGGACATTGTCTCTAAACCTCCATTAATCAATG
human_5+3_corrected   GTGGGGATGACACCAAAATGTACAAACTGGATGTGGTCTCTAAACCTCCATTAATCAATG
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA              GCCTGTATGCAAACAAGACTGTTATTAAAGCCACAGCCATTCGGCACTCCAAAAAATACT
human_5+3_corrected   GTCTGTATACAAACAGAACTGTTATTAAAGCCACAGCTGTGAGACATTCCAAAAAACACT
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA              TTGACTGCAGAGCAGATGGGATCCCATCTTCCCAGGTCACGTGGATTATGCCAGGCAATA
human_5+3_corrected   TTGACTGCAGAGCTGAAGGGACACCATCTCCTGAAGTCATGTGGATCATGCCAGACAATA
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA              TTTTCCTCCCAGCTCCATACTTTGGAAGCAGAGTCACGGTCCATCCAAATGGAACCTTGG
human_5+3_corrected   TTTTCCTCACAGCCCCCATACTATGGAAGCAGAATCACAGTCCATAAAAATGGAACCTTGG
mus_cDNA_5           ------------------------------------------------------------


rat_cDNA              AGATGAGGAACATCCGGCTTTCTGACTCTGCGGACTTCACCTGTGTGGTTCGGAGCGAGG
human_5+3_corrected   AAATTAGGAATGTGAGGCTTTCAGATTCAGCCGACTTTATCTGTGTGGCCCGAAATGAAG
mus_cDNA_5           ------------------------------------------------------------
```

```
rat_cDNA                GAGGAGAGAGTGTGTTGGTAGTGCAGTTAGAAGTCCTAGAAATGCTGAGAAGACCAACAT
human_5+3_corrected     GTGGAGAGAGCGTGTTGGTAGTACAGTTAGAAGTACTGGAAATGCTGAGAAGACCGACAT
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                TCAGAAACCCATTCAACGAAAAAGTCATCGCCCAAGCTGGCAAGCCCGTAGCACTGAACT
human_5+3_corrected     TTAGAAATCCATTTAATGAAAAAATAGTTGCCCAGCTGGGAAAGTCCACAGCATTGAATT
mus_cDNA_5             


rat_cDNA                GCTCTGTGGATGGGAACCCCCACCTGAAATTACCTGGATCTTACCTGACGGCACACAGT
human_5+3_corrected     GCTCTGTTGATGGTAACCCACCACCTGAAATAATCTGGATTTTACCAAATGGCACACGAT
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                TTGCTAACAGACCACACAATTCCCCGTATCTGATGGCAGGCAATGGCTCTCTCATCCTTT
human_5+3_corrected     TTTCCAATGGACCACAAAGTTATCAGTATCTGATAGCAAGCAATGGTTCTTTTATCATTT
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                ACAAAGCAACTCGGAACAAGTCAGGGAAGTATCGCTGTGCAGCCAGGAATAAGGTTGGCT
human_5+3_corrected     CTAAAACAACTCGGGAGGATGCAGGAAAAATATCGCTGTGCAGCTAGGAATAAAGTTGGCT
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                ACATCGAGAAACTCATCCTGTTAGAGATTGGGCAGAAGCCAGTCATTCTGACATACGAAC
human_5+3_corrected     ATATTGAGAAATTAGTCATATTAGAAATTGGCCAGAAGCCAGTTATTCTTACCTATGCAC
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                CAGGGATGGTGAAGAGCGTCAGTGGGGAACCGTTATCACTGCATTGTGTGTCTGATGGGA
human_5+3_corrected     CAGGGACAGTAAAAGGCATCAGTGGAGAATCTCTATCACTGCATTGTGTGTCTGATGGAA
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                TCCCCAAGCCAAATGTCAAGTGGACTACACCGGGTGGCCATGTAATCGACAGGCCTCAAG
human_5+3_corrected     TCCCTAAGCCAAATATCAAATGGACTATGCCAAGTGGTTATGTAGTAGACAGGCCTCAAA
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                TGGATGGAAAATACATACTGCATGAAAATGGCACGCTGGTCATCAAAGCAACAACAGCTC
human_5+3_corrected     TTAATGGGAAATACATATTGCATGACAATGGCACCTTAGTCATTAAAGAAGCAACAGCTT
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                ACGACCAAGGAAATTATATCTGTAGGGCTCAAAACAGTGTTGGCCAGGCAGTTATTAGCG
human_5+3_corrected     ATGACAGAGGAAACTATATCTGTAAGGCTCAAAATAGTGTTGGTCATACACTGATTACTG
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                TGTCAGTGATGGTTGTGGCCTACCCTCCCCGAATCATAAACTACCTACCCAGGAACATGC
human_5+3_corrected     TTCCAGTAATGATTGTAGCCTACCCTCCCCGAATTACAAATCGTCCACCCAGGAGTATTG
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                TCAGGAGGACAGGGGAAGCCATGCAGCTCCACTGTGTGTGGCCTTGGGAATCCCCAAGCCAA
human_5+3_corrected     TCACCAGGACAGGGGCAGCCTTTCAGCTCCACTGTGTGTGGCCTTGGGAGTTCCCAAGCCAG
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                AAGTCACCTGGGAGACGCCAAGACACTCCCTGCTCTCAAAAGCAACAGCAAGAAAACCCC
human_5+3_corrected     AAATCACATGGGAGATGCCTGACCACTCCCTTCTCTCAACGGCAAGTAAAGAGAGGACAC
mus_cDNA_5              ------------------------------------------------------------


rat_cDNA                ATAGAAGTGAGATGCTTCACCCACAAGGTACGCTGGTCATTCAGAATCTCCAAACCTCGG
human_5+3_corrected     ATGGAAGTGAGCAGCTTCACTTACAAGGTACCCTAGTCATTCAGAATCCCCAAACCTCCG
mus_cDNA_5              ------------------------------------------------------------
```

```
rat_cDNA              ATTCCGGAGTCTATAAGTGCAGAGCTCAGAACCTACTTGGGACTGATTACGCAACAACTT
human_5+3_corrected   ATTCTGGGATATACAAATGCACAGCAAAGAACCCACTTGGTAGTGATTATGCAGCAACGT
mus_cDNA_5            -----------------------------------------------------------


rat_cDNA              ACATCCAGGTACTCTGACAGGAAGGGGGAGACTAAAATTCAACAGAAGTCCACATCCACA
human_5+3_corrected   ATATTCAAGTAATCTGACATGAA-----ATAATAAAGT-CAACAA-----CATCTGGGCA
mus_cDNA_5           --------------------------------------------------------------


rat_cDNA              GGGTTTATTTTTTGGAAGAAGTTTAATCAAAGGCAGCCATAGGCATGTAAATGAGTCTGA
human_5+3_corrected   GAATTTATTTTTTGGAAGAAGTTTAATCAAAGGCAGCCATAGGCATGTAAATGAATTTGA
mus_cDNA_5 .         ----------------------------------------------------------


rat_cDNA              ATACATTTACAGTATTAAATTTACAATGGACATGCGA--TGA---GACTTGTAAATGAAA
human_5+3_corrected   ATACATTTACAGTATTAAATTTACAATGAACATGCAAAATAAAAGGACTTGTAAATAAAT
mus_cDNA_5           ----------------------------------------------------------


rat_cDNA              GCATTGTGAACTGAA---ACCGAGTCTCTG--TGGATCTCAAAGCAAACTCTTAACTTAA
human_5+3_corrected   GCATTATGAACTGATGATACTGATTTATTTAATGGATCTCAAAACAAACTTTTAACTTAA
mus_cDNA_5           ----------------------------------------------------------


rat_cDNA              GGCACTTTGATTTTGCCAACAAATAATAACAAACATTAAGAGAAAAAAATGATCCACTAC
human_5+3_corrected   GGCACTTTTATTTTGCCAACAAATAACAATAAACAA-----ACATTGAAACGGTTCACTAT
mus_cDNA_5           ----------------------------------------------------------


rat_cDNA              GAAATAACAAACGGCTAATGCACCTGAATTCT-CAGTAAAAAGACCTTTCTCTCGCTAAC
human_5+3_corrected   AAAATAACAAATGGCTAATGTACCTGAATTTTTCAGTAAAAAAA---TGAACTT-CTAAT
mus_cDNA_5           ----------------------------------------------------------


rat_cDNA              AGTTGCCAGCTGCCTCGTGTCTGTTTCCTACCAATGTCACAAACATCGCACACAGGGTGA
human_5+3_corrected   A-----CCAGTTGCCTAGTGTCCACCTCCTATCAATGTTACAAGCATGGCACTCAG----A
mus_cDNA_5           ----------------------------------------------------------


rat_cDNA              ATGGAGTCAACGGGAAAGATTAAGTTTGCGGTCTGTGTAAATCTCAATGTACAAATATTC
human_5+3_corrected   ACAGAGACAATGGAAAATATTAAATCTGC----------AATCT--ATGTATAAATATTT
mus_cDNA_5           ----------------------------------------------------------


rat_cDNA              TGTCNCTGGTTTATAAACATTTT-GATAAAACCGAAAAAAAAAAAAAAAAAAAAAAAAAA
human_5+3_corrected   TGT----GGTTTATAAATTTTTTGCTAAAACCTACAGAAAATAAG---------------
mus_cDNA_5           ----------------------------------------------------------


rat_cDNA              AAAA
human_5+3_corrected   ----
mus_cDNA_5           ----
```

(rat_cDNA: SEQ ID NO:7)
(human_5+3 corrected: SEQ ID NO:8)
(mus_cDNA_5: SEQ ID NO:9)

Figure 15

```
rat                MQVRGREVSGLLISLTAVCLVVTPGSRACPRRCACYVPTEVHCTFRYLTSIPDGIPANVE
human_5+3_corrected MKVKGRGITCLLVSPAVICLVATPGGKACPRRCACYMPTEVHCTFRYLTSIPDSIPPNVE
mouse_5_corrected   MQKRGREVSCLLISLTAICLVVTPGSRVCPRRCACYVPTEVHCTFRDLTSIPD-GPANVE


rat                RINLGYNSLTRLTENDFDGLSKLELIMLHSNGIHRVSDKTFSGLQSLQVLKMSYNKVQII
human_5+3_corrected RINLGYNSLVRIMETDFSGLTKLELIMLHSNGIHTIPDKTFSDLQALQVLKMSYNKVRKL
mouse_5_corrected   RVNLGYNSLTRLTENDFSGLSRLELIMLHSNGIHRVSDKTFSGLQSLQVLKMSYNKVQII


rat                RKDTFYGLGSLVRLHLDHNNIEFINPEAFYGLTSLRLVHLEGNRLTKLHPDTFVSLSYLQ
human_5+3_corrected QKDTFYGLRSLTRLHMDHNNIEFINPEVFYGLNFLRLVHLEGNQLTKLHPDTFVSLSYLQ
mouse_5_corrected   EKDTLYGLRSLTRLHLDHNNIEFINPEAFYGLTLLRLVHLEGNRLTKLHPDTFVSLSYLQ


rat                IFKTSFIKYLFLSDNFLTSLPKEMVSYMPNLESLYLHGNPWTCDCHLKWLSEWMQGNPDI
human_5+3_corrected IFKISFIKFLYLSDNFLTSLPQEMVSYMPDLDSLYLHGNPWTCDCHLKWLSDWIQ--PDV
mouse_5_corrected   IFKTSFIKXLYLYDNF-TSLPKEMVSSMPNLESLYLHGNPWTCDCHLKWLSEWMQGNP--


rat                IKCKKDRSSSSPQQCPLCMNPRISKGRPFAMVPSGAFLCTKPTIDPSLKSKSLVTQEDNG
human_5+3_corrected IKCKKDRSPSSAQQCPLCMNPRTSKGKPLAMVSAAAFQCAKPTIDSSLKSKSLTILEDSS
mouse_5_corrected   ------------------------------------------------------------


rat                SASTSPQDFIEPFGSLSLNMTXXSGNKADMVCSIQKPSRTSPTAFTEENDYIMLNASPST
human_5+3_corrected SAFISPQGFMAPFGSLTLNMTDQSGNEARMVCSIQKPSRTSPIAFTEENDYIVLNTSFST
mouse_5_corrected   ------------------------------------------------------------


rat                NLVCSVDYNHIQPVWQLLALYSDSPLILERKPQLTETPSLSSRYKQVALRPEDIFTSIEA
human_5+3_corrected FLVCNIDYGHIQPVWQILALYSDSPLILERSHLLSETPQLYYKYKQVAPKPEDIFTNIEA
mouse_5_corrected   ------------------------------------------------------------


rat                DVRADPFWFQQEKIVLQINRTATTLSTLQIQFSTDAQIALPRAEMRAERLKWTMILMMNN
human_5+3_corrected DLRADPSWLMQDQISLQLNRTATTFSTLQIQYSSDAQITLPRAEMRPVKHKWTMISRDNN
mouse_5_corrected   ------------------------------------------------------------


rat                PKLERTVLVGGTIALSCPGKGDPSPHLENLIADGSKVRAPYVSEDGRILIDKNGKLELQM
human_5+3_corrected TKLEHTVLVGGTVGLNCPGQGDPTPHVDWLLADGSKVRAPYVSEDGRILIDKSGKLELQM
mouse_5_corrected   ------------------------------------------------------------


rat                ADSFDAGLYHCISTNDADADVLTYRITVVEPYGESTHDSGVQHTVVTGETLDLPCLSTGV
human_5+3_corrected ADSFDTGVYHCISSNYDDADILTYRITVVEPLVEAYQENGIHHTVFIGETLDLPCHSTGI
mouse_5_corrected   ------------------------------------------------------------


rat                PDASISWILPGNTVFSQPSRDRQILNNGTLRILQVTPKDQGHYQCVAANPSGADFSSFKV
human_5+3_corrected PDASISWVIPGNNVLYQSSRDKKVLNNGTLRILQVTPKDQGYYRCVAANPSGVDFLIFQV
mouse_5_corrected   ------------------------------------------------------------


rat                SVQKKGQRMVEHDREAGGSGLGEPNSSVSLKQPASLKLSASALTGSEAGKQVSGVHRKNK
human_5+3_corrected SVKMKGQRPLEHDGETEGSGLDESNPIAHLKEPPGAQLRTSALMEAEVGKHTSSTSKRHN
mouse_5_corrected   ------------------------------------------------------------


rat                HRDLIHRRRGDSTLRRFREHRRQLPLSARRIDPQRWAALLEKAKKNSVPKKQENTTVKPV
human_5+3_corrected YRELTLQRRGDSTHRRFRENRRHFPPSARRIDPQHWAALLEKAKKNAMPDKRENTTVSPP
mouse_5_corrected   ------------------------------------------------------------
```

```
rat                 PLAVPLVELTDEEKDASGMIPPDEEFMVLKTKASGVPGRSPTADSGPVNHGFMTSIASGT
human_5+3_corrected PVVTQLPNIPGEEDDSSGMIALHEEFMVPATKAINLPARTVTADSRTISDSPMTNINYGT
mouse_5_corrected   ------------------------------------------------------------


rat                 EVS-TVNPQTLQSEHLPDFKLFSVTNGTAVTKSMNPSIASKIEDTTNQNPIIIFP----SV
human_5+3_corrected EFSPVVNSQILPPEEPTDFKLSTAIKTTAMSKNINPTMSSQIQGTTNQHSSTVFPLLLGA
mouse_5_corrected   ------------------------------------------------------------


rat                 AEIRDSAQAGRAS--SQSAHPVTGGNMATYGHTNTYSSFTSKASTVLQPINPTESYGPQI
human_5+3_corrected TEFQDSDQMGRGREHFQSRPPITVRTMIKDVNVKMLSSTTNKL--LLESVNTTNSH--QT
mouse_5_corrected   ------------------------------------------------------------


rat                 PITGVSRPSSSDISSHTTADPSFSSHPSGSHTTASSLFHIPRNNNTGNFPLSRHLGRERT
human_5+3_corrected SVREVSEPRENHFYSHTTQILSTSTFPSDPHTAAHSQFPIPRNS-TVNIPLFRRFGRQRK
mouse_5_corrected   ------------------------------------------------------------


rat                 IWSRGRVKNPHRTPVLRRHRHRTVRPAIKGPANKNVSQVPATEYPGMCHTCPSAEGLTVA
human_5+3_corrected IGGRGRIISPYRTPVLRBHRYSIFRSTTRGSSEKSTTAFSATVLNVTCLSCLPRERLTTA
mouse_5_corrected   ------------------------------------------------------------


rat                 TAALSVPSSSHSALPKTNNVGVIAEESTTVVKKPLLLFKDKQNVDIEIITTTTKYSGGES
human_5+3_corrected TAALSFPSAAPITFPKADIARVPSEESTTLVQNPLLLLENKP--SVEKTTPTIKYFRTEI
mouse_5_corrected   ------------------------------------------------------------


rat                 NHVIPTEASMTSAPTSVSLGKSPVDNSGHLSMPGTIQTGKDSVETTPLPSPLSTP--SIP
human_5+3_corrected SQVTPTGAVMTYAPTSIPMEKTHKVNASYPRVSSTMEAKRDSVITSSLSGAITKPPMTII
mouse_5_corrected   ------------------------------------------------------------


rat                 TSTKFSKRKTPLHQIFVNNQKKEGMLKNPYQFGLQKNPAAKLPKIAPLLPTGQSSPSDST
human_5+3_corrected AITRFSRRKIPWQQNFVNNHNPKGRLRNQHKVSLQKSTAVMLPKTSPALPQRQSSPFHFT
mouse_5_corrected   ------------------------------------------------------------


rat                 TLLTSPPPALSTTMAATQNKGTEVVSGARSLSAGKKQ-PFTNSSPVLPSTISKRSNTLNF
human_5+3_corrected TLSTSVMQIPSNTLTTAHHTTTKTHNPG-SLPT-KKELPFPPLNPMLPSIISKDSSTKSI
mouse_5_corrected   ------------------------------------------------------------


rat                 LSTETPT-VTSPTATASVIMSETQRTRSKEAKDQIKG-P-RKNRNNANTTPRQVSGYSAY
human_5+3_corrected ISTQTAIPATTPTFPASVITYETQTERSRAQTIQREQEPQKKNRTDPNISPDQSSGFTTP
mouse_5_corrected   ------------------------------------------------------------


rat                 SALTTADTPLAFSHSPRQDDGGNVSAVAYHSTTS--LLAITELFEKYTQTLGNTTALETT
human_5+3_corrected TAMTP---PALAFTHSPPENTTGISSTISFHSRTLNLTDVIEELAQASTQTLKSTIASETT
mouse_5_corrected   ------------------------------------------------------------


rat                 LLSKSQESTTVKRAS-DTP-PFLLSSGAPPVPTPSPPPFTKGVVTDSKVTSAFQMTSNRV
human_5+3_corrected LSSKSHQSTTFRKASLDTPIPPFLSSSATIMPVPISPPFTQRAVTDTRGDSHFRIMTNTV
mouse_5_corrected   ------------------------------------------------------------


rat                 VTIYESSRHNTDLQQPSAEASPNPEIITGTTDSPSNLFPSTSVPALRVDKPQNSKWKPSP
human_5+3_corrected VKLHESSRHN--LQMPSSQLEP---------ITSSTSNLLHSTPMPALTTVKSQNSKLTPSP
mouse_5_corrected   ------------------------------------------------------------


rat                 WPEHKYQLKSYSETIERGKRPAVSMSPHLSLPEASTHASHWNTQKHAEKSVFDKKPGQNP
human_5+3_corrected WAEYQFWHKPYSDIAEKGKKPEVSMIATTGLSEATTLVSDWDGQRNTKKSDFDKKPVQEA
mouse_5_corrected   ------------------------------------------------------------
```

```
rat                  -TSKHLPYVSLPKTLLKKPRIIGGKAASFTVPANSDVFLPCEAVGDPLPIIHWTRVSSGX
human_5+3_corrected  TTSKLLPFDSLSRYIFEKPRIVGGKAASFTIPANSDAFLPCEAVGNPLPTIHWTRVS-GL
mouse_5_corrected    ------------------------------------------------------------


rat                  KISQGTQKSRFHVLPNGTLSIQRVSIQDRGQYLCSAFNPLGVDHFHVSLSVVFYPARILD
human_5+3_corrected  DLSRGNQNSRVQVLPNGTLSIQRVEIQDRGQYLCSASNLFGTDHLHVTLSVVSYPPRILE
mouse_5_corrected    ------------------------------------------------------------


rat                  RHVKEITVHFGSTVELKCRVEGMPRPTVSWIIANQTVVSETAKGSRKVWVTPDGTLIIYN
human_5+3_corrected  RRTKEITVHSGSTVELKCRAEGRPSPTVTWIIANQTVVSESSQGSRQAVVTVDGTLVLHN
mouse_5_corrected    ------------------------------------------------------------


rat                  LSLYDRGFYKCVASNPSGQDSLLVKIQVITAPPVIIEQKROAIVGVLGGSLKLPCTAKGT
human_5+3_corrected  LSIYDRGFYKCVASNPGGQDSLLVKIQVIAAPPVILEQRRQVIVGTWGESLKLPCTAKGT
mouse_5_corrected    ------------------------------------------------------------


rat                  PQPSVHWVLYDGTELKPLQLTHSRFFLYPNGTLYIRSIAPSVRGIYECIATSSSGSERRV
human_5+3_corrected  PQPSVYWVLSDGTEVKPLQFTNSKLFLFSNGTLYIRNLASSDRGTYECIATSSTGSERRV
mouse_5_corrected    ------------------------------------------------------------


rat                  VILTVEEGETIPRIETASQKWTEVNLGEKLLLNCSATGDPKPRIIWRLPSKAVIDQWHRM
human_5+3_corrected  VMLTMEERVTSPRIEAASQKRTEVNFGDKLLLNCSATGEPKPQIMWRLPSKAVVDQ----
mouse_5_corrected    ------------------------------------------------------------


rat                  GSRIHVYPNGSLVVGSVTEKDAGDYLCVARNKMGDDLVLMHVRIRLTPAKIEQKQYFKKQ
human_5+3_corrected  GSWIHVYPNGSLFIGSVTEKDSGVYLCVARNKMGDDLILMHVSLRLKPAKIDHKQYFRKQ
mouse_5_corrected    ------------------------------------------------------------


rat                  VLHGKDFQVDCKASGSPVPEVSWSLPDGTVLNNVAQADDSGYRTKRYTLFHNGTLYFNNV
human_5+3_corrected  VLHGKDFQVDCKASGSPVPEISWSLPDGTNINNAMQADDSGHRTRRYTLFNNGTLYFNKV
mouse_5_corrected    ------------------------------------------------------------


rat                  GMAEEGDYICSAQNTLGKDEMKVHLTVLTAIPRIRQSYKTTMRLRAGETAVLDCEVTGEP
human_5+3_corrected  GVAEEGDYTCYAQNTLGKDEMKVHLTVITAAPRIRQSNKTNKRIKAGDTAVLDCEVTGDP
mouse_5_corrected    ------------------------------------------------------------


rat                  KPNVFWLLPSNNVISFSNDRFTPHANRTLSIHKVKPLDSGDYVCVAQNPSGDDTKTYKLD
human_5+3_corrected  KPKIFWLLPSNDMISFSIDRYTFHANGSLTINKVKLLDSGEYVCVARNPSGDDTKMYKLD
mouse_5_corrected    ------------------------------------------------------------


rat                  IVSKPPLINGLYANKTVIKATAIRHSKKYFDCRADGIPSSQVTWIMPGNIFLPAPYFGSR
human_5+3_corrected  VVSKPPLINGLYTNRTVIKATAVRHSKKHFDCRAEGTPSPEVMWIMPDNIFLTAPYYGSR
mouse_5_corrected    ------------------------------------------------------------


rat                  VTVHPNGTLEMRNIRLSDSADFTCVVRSEGGESVLVVQLEVLEMLRRPTFRNPFNEKVIA
human_5+3_corrected  ITVHKNGTLEIRNVRLSDSADFICVARNEGGESVLVVQLEVLEMLRRPTFRNPFNEKIVA
mouse_5_corrected    ------------------------------------------------------------


rat                  QAGKPVALNCSVDGNPPPEITWILPDGTQFANRPHNSPYIMAGNGSLILYKATRNKSGKY
human_5+3_corrected  QLGKSTALNCSVDGNPPPEIIWILPNGTRFSNGPQSYQYLIASNGSFIISKTTREDAGKY
mouse_5_corrected    ------------------------------------------------------------


rat                  RCAARNKVGYIEKLILLEIGQKPVILTYEPGMVKSVSGEPLSLHCVSDGIPKPNVKWTTP
human_5+3_corrected  RCAARNKVGYIEKLVTLEIGQKPVILTYAPGTVKGISGESLSLHCVSDGIPKPNIKWTMP
mouse_5_corrected    ------------------------------------------------------------
```

```
rat                   GGHVIDRPQVDGKYILHENGTLVIKATTAHDQGNYICRAQNSVGQAVIEVSVMVVAIPPK
human_5+3_corrected   SGYVVDRPQINGKYILHDNGTLVIKEATAYDRGNYICKAQNSVGETLITVPVMIVAYPPR
mouse_5_corrected     ------------------------------------------------------------


rat                   IINYLPRNMLRRTGEAMQLHCVALGIPKPKVTWETPRHSLLSKATARKPHRSEMLHPQGT
human_5+3_corrected   ITNRPPRSIVTRTGAAFQLHCVALGVPKPEITWEMPDHSLLSTASKERTHGSEQLHLQGT
mouse_5_corrected     ------------------------------------------------------------


rat                   LVIQNLQTSDSGVYKCRAQNLLGTDYATTYIQV
human_5+3_corrected   LVIQNPQTSDSGIYKCTAKNPLGSDYAATYIQV
mouse_5_corrected     ---------------------------------
```

(rat: SEQ ID NO:10)
(human_5+3_corrected: SEQ ID NO:11)
(mouse_5_corrected: SEQ ID NO:12)

Figure 16

```
rat                MQVRGREVSGLLISLTAVCLVVTPGSRACPRRCACYVPTEVECTFRYLTSIPDGIPANVE
human_5+3_corrected MKVKGFGITCLLVSFAVICLVATPGGKACPRRCACYMPTEVHCTFRYLTSIPDSIPPNVE
                   *:*:**  ::  **:*:_.:***.***.;*********;*;******+******.**.***

rat                RINLGYNSLTRLTENDFDGLSKLELIMLHSNGIHRVSDKTFSGLQSLQVLKMSYNKVQII
human_5+3_corrected RINLGYNSLVRIMETDFSGLTKLELIMLHSNGIHTIPDKTFSDLQALQVLKMSYNKVRKL
                   *********.**  *_.**.**;:*****+*****  ;_.*****.**;**********; :

rat                RKDTFYGLGSLVRLHLDHNNIEFINPEAFYGLTSLRLVHLEGNRLTKLHPDTFVSLSYLQ
human_5+3_corrected QKDTFYGLRSLTRLHMDHNNIEFINPEVFYGLNFLRLVHLEGNQLTKLHPDTFVSLSYLQ
                   :*******  **.**;************+  ;****.  *********;***********:*****

rat                IFKTSFIKYLFLSDNFLTSLPKEMVSYMPNLESLYLHGNPWTCDCHLKWLSEWMQGNPDI
human_5+3_corrected IFKISFIKFLYLSDNFLTSLPQEMVSYMPDLDSLYLHGNPWTCDCHLKWLSDWIQ--PDV
                   ***  ****;*;**********;******;*;***********************:*:* **;

rat                IKCKKDRSSSSPQQCPLCMNPRISKGRPFAMVPSGAFLCTKPTIDPSLKSKSLVTQEDNG
human_5+3_corrected IKCKKDRSPSSAQQCPLCMNPRTSKGKPLAMVSAAAFQCAKPTIDSSLKSKSLTILEDSS
                   *********.**.**********  ***:*:***_.:.**  *:*****.*******.  **..

rat                SASTSPQDFIEPFGSLSIAMTKXSGNKADMVCSIQKPSRTSPTAFTEENDYIMLXASFST
human_5+3_corrected SAFISPQGFMAPFGSLTLNMTDQSGNEANMVCSIQKPSRTSPIAFTEENDYIVLNTSFST
                   **  ***.*: *****;**** ***.*;***********+ *********;**;:****

rat                NLVCSVDYNHIQPVWQLLALYSDSPLILERKPQLTETPSLSSRYKQVALRPEDIFTSIEA
human_5+3_corrected FLVCNIDYGHIQPVWQIIALYSDSPLILERSHLLSETPQLYYKYKQVAPKPEDIFTNIEA
                   ***.:**.*******;**********;***.*  *;***.*  ;***** ;******.***

rat                DVRADPFWFQQEKIVLQLNRTATTLSTLQIQFSTDAQIALPRAEMRAERLKWTMILMNNN
human_5+3_corrected DLRADPSWLMQDQISLQLNRTATTFSTLQIQYSSDAQITLPRAEMRPVKHKWTMISRDNN
                   *:****  *: *::*  ************;*****;*;****;*******.  ; ***** **

rat                PKLERTVLVGGTIALSCPGKGDPSPHLEWLLADGSKVRAPYVSEDGRILIDKNGKLELQM
human_5+3_corrected TKLEHTVLVGGTVGLNCPGQGDPTPHVDWLLADGSKVRAPYVSEDGRILIDKSGKLELQM
                   .***;*********;.*.***;***;***;*;;*********************.******;*

rat                ADSFDAGLYRCISTNDADADVLTYRITVVEPYGESTHDSGVQHTVVTGETLDLPCLSTGV
human_5+3_corrected ADSFDTGVYRCISSNYDDADILTYRITVVEPLVEAYQENGIHHTVFIGETLDLPCHSTGI
                   *****;*;*****;*  ***;**********  *;  ::.*;::*** ;******* ***;

rat                PDASISWILPGNTVFSQPSRDKQILANGTLRILQVTPKDQGHYQCVAANPSGADFSSFKV
human_5+3_corrected PDASISWVIPGNNVLYQSSRDKKVLNNGTLRILQVTPKDQGYIRCVAANPSGVDFLIFQV
                   *******;;***.*;  *.***:;::**************;;********** ***  *;*

rat                SVQKKGQRMVEHDREAGGSGLGEPNSSVSLKQPASLKLSASALTGSEAGKQVSGVHRKNK
human_5+3_corrected SVKMKGQRPLEHDGETEGSGLDESNPIAHLKEPPGAQLRTSALMEAEVGKHTSSTSKRHN
                   **:;****  ;*** *; ****.*.*.  . **;*..  ;* ;***  ;*.**;.*.. ::::

rat                HRDLIHRRRGDSTLRRFREHRRQLPLSARRIDPQRWAALLEKAKKNSVPKKQENTTVKPV
human_5+3_corrected YRELTLQRRGDSTHRRFRENRRHFPPSARRIDPQHWAALLEKAKKNAMPDKRENTTVSPP
                   :*;*  ;******  ****;**;:*;* *********;**********;*.*;:*****.*

rat                PIAVPLVELTDEEKDASGMIPPDEEFHVLKTKASGVPGRSPTADSGPVNHGFMTSIASGT
human_5+3_corrected PVVTQLPNIPGEEDSSGMLALHEEFNVPATKALNLPARIVTADSRTISDSPMTNINYGT
                   *:..  * ::..**.*;*****:.  .*****  **  .;*.*;  ****  .;... **.*  **

rat                EVS-TVNPQTLQSEHLPDFKLFSVTNGTAVTKSMNPSIASKIEDTTNQNPIIIFP---SV
human_5+3_corrected EFSPVVNSQILPPEEPTDFKLSTAIKTTAMSKNINPTMSSQIQGTTNQHSSTVFPLLLGA
                   *.*  .**.*  *  .*.  .****  ::. : **::*.:**:::*:*:.****;.  ;**  ..

rat                AEIRDSAQAGRAS--SQSAHPVTGGNMATYGHTNTYSSFTSKASTVLQPINPTESYGPQI
human_5+3_corrected TEFQDSDQMGRGREHFQSRPPITVRTMIKDVNVKMLSSTTNKL--LLESVNTTNSR--QT
                   :*;:;**  *  r*.     **  *;* .* .   :.;  **  *.*    :*:.;*.*;:*; *
```

```
rat                   PITGVSRPSSSDISSHTTADPSFSSHPSGSHTTASSLFHIPRNNNTGNFPLSRHLGRERT
human_5+3_corrected   SVREVSEPRRNHFYSHTTQILSTSTFPSDPHTAAHSQFPIPRNS-TVNIPLFRRFGRQRK
                      .: **.*  ..: ****   * *:.**..**:.* * * ****. * *:** *:;**:*.

rat                   IWSRGRVKNPHRTPVLRRHRHRTVRPAIKGPANKNVSQVPATEYPGMCHTCPSAEGLTVA
human_5+3_corrected   IGGRGRIISPYRTPVLRRHRYSIFRSTTRGSSEKSTTAFSATVLNVTCLSCLPRERLTTA
                      *  .***: .*:********** .*.: :*,::*..: ..** * :* .* ** *

rat                   TAALSVPSSSHSALPKTNNVGVIAEESTTVVKKPLLLFKDKQNVDIEIITTTTKYSGGES
human_5+3_corrected   TAALSFPSAAPITFPKADIARVPSEESTTLVQNPLLLLENKP--SVEKTTPTIKYFRTEI
                      *****.**:: ::*:*:: . * ;*****:*;:***::;* .:* *.* ** *

rat                   NHVIPTEASMTSAPTSVSLGKSPVDNSGHLSMPGTIQTGKDSVETTPLPSPLSTP--SIP
human_5+3_corrected   SQVTPTGAVMTYAPTSIPMEKTHKVNASYPRVSSTNEAKRDSVITSSLSGAITKPPMTII
                      .:* ** * ** *****;.: *;   *:.: :,.* :: :*** *:.*...::.* :*

rat                   TSTKFSKRKTPLHQIFVNNQKKEGMLKNPYQEGLQKNPAAKLPKIAPLLPTGQSSPSDST
human_5+3_corrected   AITRFSRRKIPWQQNFVNNHNPKGRLRNQHKVSLQKSTAVMLPKTSPALPQRQSSPFHFT
                      ; *:**;** * :* ****:; :* *:* ;1,.***..*. *** :* ** **** . *

rat                   TLLTSPPPALSTTMAATQNKGTEVVSGARSLSAGKKQPFTNSSPVLPSTISKRSNTLNFL
human_5+3_corrected   TLSTSVMQIPSNTLTTAHHTTTKTHNPG-SLPTKKELPFPPLNPMLPSIISKDSSTKSII
                      ** **     *.*::::11, *:. . . **.: *: **. .*:*** *** *.* .::

rat                   STETPT-VTSPTATASVIMSETQRTRSKEARIQIKG-P-RKNRNNANTTPRQVSGYSAYS
human_5+3_corrected   STQTAIPATTPTFPASVITYETQTERSRAQTIQREQEPQKKQRTDPNISPDQSSGFTTPT
                      **:*. .*:** .**** *** **; . * : ;***.:.* :* * **::: :

rat                   ALTTADTPLAFSHSPRQDDGGNVSAVAYHSTTS--LLAITELFEKYTQTLGNTTALETTL
human_5+3_corrected   AMTP---PALAFTHSPPENTTGISSTISFHSRTLNLTDVIEELAQASTQTLKSTIASETTL
                      *:*. ..***;*** :: * *;;;;** * .* ** : **** .* * ****

rat                   LSKSQESTTVKRAS-DTP-PPLLSSGAPPVPTPSPPPFTKGVVTDSKVTSAFQMTSNRVV
human_5+3_corrected   SSKSHQSTTTRKASLDTPIPPFLSSSATLMPVPISPPPFTQRAVTDTRGDSRFRIMTNTVV
                      ***::***.;;** *** **;***.*. :*.* .****: .***:: * *;; ;* **

rat                   TIYESSRHNTDLQQPSAEASPNPEIITGTTDSPSNLFPSTSVPALRVDKPQNSKWKPSPW
human_5+3_corrected   KLHESSRHN--LQMPSSQLEP-------LTSSTSNLLHSTPMPALTTVKSQNSKLTPSPW
                      .::******  ** **:: .*     *.*.***: **.:*** . *.**** .****

rat                   PEHKYQLKSYSETIEKGKRPAVSMSPHLSLPEASTHASHWNTQKHAEKSVFDKKPGQNP-
human_5+3_corrected   AEYQFWHKPYSDIAEKGKKPEVSMLATTGLSEATTLVSDWDGQKNTKKSDFDKKPVQEAT
                      .*;;: *.**: ****:* *** , .*.**:*_.*.*; **:;;** ***** *:.

rat                   TSKHLPYVSLPKTLLKKPRIIGGKAASFTVPANSDVFLPCEAVGDPLPIIHWTRVSSGXE
human_5+3_corrected   TSKLLPFDSLSRYIFEKPRIVGGKAASFTIPANSDAFLPCEAVGNPLPTIHWTRVS-GLD
                      *** **; **.; :::****;********* ;**** .********;*** ******* * :

rat                   ISQGTQKSRFHVLPNGTLSIQRVSIQDRGQYLCSAFNPLGVDHFHVSLSVVFYPARILDR
human_5+3_corrected   LSRGNQNSRVQVLPNGTLSIQRVEIQDRGQYLCSASNLFGTDHLHVTLSVVSYPPRILER
                      :*;*.*;**.;*********** .********* * :*.**;**;**** **.***;*

rat                   HVKEITVHFGSTVELKCRVEGMPRPTVSWILANQTVVSETAKGSRKVWVTPDGTLIIYNL
human_5+3_corrected   RTKEITVHSGSTVELKCRAEGRPSPTVTWILANQTVVSESSQGSRQAVVTVDGTLVLHNL
                      :.****** *********.** * ***;************;;;***:. ** ****::;**

rat                   SLYDRGFYKCVASNPSGQDSLLVKIQVITAPPVIIEQKRQAIVGVLGGSLKLPCTAKGTP
human_5+3_corrected   SIYDRGFYKCVASNPGGQDSLLVKIQVIAAPPVILEQRRQVIVGTWGESLKLPCTAKGTP
                      *;***********;***.***********:****:**.**:**.** * ************

rat                   QPSVHWVLYDGTELKPLQLTHSRFFLYPNGTLYIRSIAPSVRGTYECIATSSSGSERRVV
human_5+3_corrected   QPSVYWVLSDGTEVKPLQFTNSKLFLFSNGTLYIRNLASSDRGTYECIATSSTGSERRVV
                      ****;*** ****;****;*:;:**;.******;** .:* * ************;*******

rat                   ILTVEEGETIPRIETASQKWTEVNLGEKLLNCSATGDPKPRTIWRLPSKAVIDQWHRMG
human_5+3_corrected   MLTMEERVTSPRIEAASQKRTEVNFGDKLLLNCSATGEPKPQIMWRLPSKAVVDQ-----G
                      ;**:** * ****:**** *:**:*:***********:***: :*************.*:**     *
```

```
rat                SRIHVYPNGSLVVGSVTEKDAGDYLCVARNKMGDDLVIMHVRLRLTPAKIEQKQYFKKQV
human_5+3_corrected SWIHVYPNGSLFIGSVTEKDSGVYLCVARNKMGDDLIIMHVSLRLKPAKIDHKQYFRKQV
                   * ********** ;******;* ****************;**** ***.****;;****;***

rat                LHGKDFQVDCKASGSPVPEVSWSLPDGTVLNNVAQADDSGYRTKRYTLFHNGTLYFNNVG
human_5+3_corrected LHGKDFQVDCKASGSPVPEISWSLPDGTMINNAMQADDSGHRTRRYTLFNNGTLYFNKVG
                   ******************;***;*****;;**. *******;**;*****;*******;**

rat                MAEEGDYICSAQNTLGKDEMKVHLTVLTAIPRIRQSYKTTMRLRAGETAVLDCEVTGEPK
human_5+3_corrected VAEEGDYTCYAQNTLGKDEMKVHLTVITAAPRIRQSNKTNKRIKAGDTAVLDCEVTGDPK
                   ;****** * *****************;** *****;* **. *;;**;***********;**

rat                PNVFWLLPSMNVISFSNDRFTFHANRTLSIHKVKPLDSGDYVCVAQNPSGDDTRTYKLDI
human_5+3_corrected PKIFWLLPSNDMISFSIDRYTFHANGSLTINKVKLLDSGEYVCVARNPSGDDTKMYKLDV
                   *;;*******;;**** **;***** ;*;*;*** ****;****;******** ****;

rat                VSKPPLINGLYANKTVIKATAIRHSKKYFDCRADGIPSSQVTWIMPGNIFLPAPYFGSRV
human_5+3_corrected VSKPPLINGLYTNRTVIKATAVRHSKKHFDCRAEGTPSPEVMWIMPDNIFLTAPYYGSRI
                   ***********;*;*******;*****;*****;* **.;* **** **** ***;***;

rat                TVHPNGTLEMRNIRLSDSADFTCVVRSEGGESVLVVQLEVLEMLRRPTFRNPFNEKVIAQ
human_5+3_corrected TVHKNGTLEIRNVRLSDSADFICVARNEGGESVLVVQLEVLEMLRRPTFRNPFNEKIVAQ
                   *** ****;**;**;******* **.*.**********************;;;**

rat                AGKPVALNCSVDGNPPPEITWILPDGTQFANRPHNSPYLMAGNGSLILYKATRNKSGKYR
human_5+3_corrected LGKSTALNCSVDGNPPPEIIWILPNGTRFSNGPQSYQYLIASNGSFIISKTTREDAGKYR
                    **..**************;****;**;*;* *;. **;*.***;*; *;**:..;****

rat                CAARNKVGYIEKLILLEIGQKPVILTYEPGMVKSVSGEPLSLHCVSDGIPKPNVKWTTPG
human_5+3_corrected CAARNKVGYIEKLVILEIGQKPVILTYAPGTVKGISGESLSLHCVSDGIPKPNIKWTMPS
                   *************;;************ ** **.;*** *************;*** *.

rat                GHVIDRPQVDGKYILHENGTLVIKATTAHDQGNYICRAQNSVGQAVISVSVMVVAYPPRI
human_5+3_corrected GYVVDRPQINGKYILHDNGTLVIKEATAYDRGNYICKAQNSVGHTLITVPVMIVAYPPRI
                   *;*;****;;****** ;******* ;**;*;***** ***** ;;;*;*.**;*******

rat                INYLPRNMLRRTGEAMQLHCVALGIPKPKVTWETPRHSLLSKATARKPHRSEMLHPQGTL
human_5+3_corrected TNRPPRSIVTRTGAAFQLHCVALGVPKPEITWEMPDHSLLSTASKERTHGSEQLHLQGTL
                    *  **.;; *** *;********;***;;**** * *****.*; ..;.* ** ** ****

rat                VIQNLQTSDSGVYKCRAQNLLGTDYATTYIQVL
human_5+3_corrected VIQNPQTSDSGIYKCTAKNPLGSDYAATYIQVI
                   **** ******;*** *;* **;***;**;***;
```

(rat: SEQ ID NO:13)

(human_5+3_corrected: SEQ ID NO:14)

Figure 17

MQKRGREVSCLLISLTAICLVVTPGSRVCPRRCACYVPTEVHCTFRDLTSIPDGPANVER
VNLGYNSLTRLTENDFSGLSRLELLMLHSNGIHRVSDKTFSGLQSLQVLKMSYNKVQIIE
KDTLYGLRSLTRLHLDHNNIEFINPEAFYGLTLLRLVHLEGNRLTKLHPDTFVSLSYLQIF
KTSFIKXLYLYDNFTSLPKEMVSSMPNLESLYLHGNPWTCDCHLKWLSEWMQGNP
(SEQ ID NO: 15)

Figure 18

MKVKGRGITCLLVSFAVICLVATPGGKACPRRCACYMPTEVHCTFRYLTSIPDSIPPNVE
RINLGYNSLVRLMETDFSGLTKLELLMLHSNGIHTIPDKTFSDLQALQVLKMSYNKVRK
LQKDTFYGLRSLTRLHMDHNNIEFINPEVFYGLNFLRLVHLEGNQLTKLHPDTFVSLSYL
QIFKISFIKFLYLSDNFLTSLPQEMSYMPDLDSLLYLHGNPWTCDCHLRWLSDWIQPDVI
KCKKDRSPSSAQQCPLCMNPRTSKGKPLAMVSAAAFQCAKPTIDSSLKSKSLTILEDSSS
AFISPQGFMAPFGSLTLNMTDQSGNEANMVCSIQKPSRTSPIAFTEENDYIVLNTSFSTFL
VCNIDYGHIQPVWQILALYSDSPLILERSHLLSETPQLYYKYKQVAPKPEDIFTNIEADLR
ADPSWLMQDQISLQLNRTATTFSTLQIQYSSDAQITLPRAEMRPVKHKWTMISRDNNTK
LEHTVLVGGTVGLNCPGQGDPTPHVDWLLADGSKVRAPYVSEDGRILIDKSGKLELQM
ADSFDTGVYHCISSNYDDADILTYRITVVEPLVEAYQENGIHHTVFIGETLDLPCHSTGIP
DASISWVIPGNNVLYQSSRDKKVLNNGTLRILQVTPKDQGYYRCVAANPSGVDFLIFQV
SVKMKGQRPLEHDGETEGSGLDESNPIAHLKEPPGAQLRTSALMEAEVGKHTSSTSKRH
NYRELTLQRRGDSTHRRFRENRRHFPPSARRIDPQHWAALLEKAKKNAMPDKRENTTV
SPPPVVTQLPNIPGEEDDSSGMLALHEEFMVPATKALNLPARTVTADSRTISDSPMTNIN
YGTEFSPVVNSQILPPEEPTDFKLSTAILTTAMSKNINPTMSSQIQGTTNQHSSTVFPLLLG
ATEFQDSDQMGRGREHFQSRPPITVRTMIKDVNVKMLSSTTNKLLLESVNSHQTSVREV
SEPRHNHFYSHTTQILSTSTFPSDPHTAAHSQFPIPRNSTVNIPLFRRFGRQRKIGGRGRIIS
PYRTPVLRRHRYSIFRSTTRGSSEKSTTAFSATVLNVTCLSCLPRELTTATAALSFPSAAPI
TFPKADIARVPSEESTTLVQNPLLLLENKPSVEKTTPTIKYFRTEISQVTPTGAVMTYAPT
SIPMEKTHKVNASYPRVSSTNEAKRDSVITSSLSGAITKPPMTIIAITRFSRRKIPWQQNFV
NNHNPKGRLRNQHKVSLQKSTAVMLPKTSPALPQRQSSPFHFRRLSTSVMQIPSNTLTT
AHHTTTKTHNPGSLPTKKELPFPPLNPMLPSIISKDSSTKSIISTQTAIPATTPTFPASVITYE
TQTERSRAQTIQREQEPQKKNRTDPNISPDQSSGFTTPTAMTPPALAFTHSPPENTTGISST
ISFHSRTLNLTDVIEELAQASTQTLKSTIASETTLSSKSHQSTTTRKASLDTPIPPFLSSSAT
LMPVPISPPFTQRAVTDTRGDSHFRLMTNTVVKLHESSRHNLQMPSSQLEPLTSSTSNLL
HSTPMPALTTVKSQNSKLTPSPWAEQFWHKPYSDIAEKGKKPEVSMLATTGLSEATTLV
SDWDGQKNTKKSDFDKKPVQEATTSKLLPFDSLSRYIFEKPRIVGGKAASFTIPANSDAF
LPCEAVGNPLPTIHWTRVSGLDLSRGNQNSRVQVLPNGTLSIQRVEIQDRFQYLCSASNL
FGTDHLHVTLSVVSYPPRILERRTKEITVHSGSTVELKCRAEGRPSPTVTWILANQTVVSE
SSQGSRQAVVTVDGTLVLHNLSIYDRGFYKCVASNPGGQDSLLVKIQVIAAPPVILEQRR
QVIVGTWGESLKLPCTAKGTPQPSVYWVLSDGTEVKPLQFTNSKLFLFSNGTLYIRNLAS

SDRFTYECIATSSTGSERRVVMLTMEERVTSPRIEAASQKRTEVNFGDKLLLNCSATGEP
KPQIMRLPSKAVVDQGSWIHYPNGSLFIGVTEKDSGVYLCVARNKMGDDLILMHVSLR
LKPAKIDHKQYFRKQVLHGKDFQVDCKASGSPVPEISWSLPDGTMINNAMQADDSGHR
TRRYRLFNNGTLYFNKVGVAEEGDYTCYAQNTLGKDEMKVHLTVITAAPRIRQSNKTN
KRIKAGDTAAVLDCEVTGDPKPKIFWLLPSNDMISFSIDRYTFHANGSLTINKVKLLDSG
EYVCVARNPSGDDTKMYKLDVVSKPPLINGLYTNRTVIKATAVRHSKKHFDCRAEGTP
SPEVMWIMPDNIFLTAPYYGSRITVHKNGTLEIRNVRLSADFICVARNEGGESVLVVQLE
VLEMLRRPTFRNPRNPFNEKIVAQLGKSTALNCVSVDGNPPPEIIWILPNGTRFSNGPQSY
QYLIASNGSFIISKTTREDAGKYRCAARNKVGYIEKLVILEIGQKPVILTYAPGTVKGISGE
SLSLHCVSDGIPKPNIKWTMPSGYVVDRPQINGKYILHDNGTLVIKEATAYDRGNYICKA
QNSVGHTLITVPVMIVAYPPRITNRPPRSIVTRTGAAFQLHCVALGVPKPEITWEMPDHS
LLSTASKERTHGSEQLHLQGTLVIQNPQTSDSGIYKCTAKNPLGSDYAATYIQVI (SEQ ID
NO: 16)

Figure 19
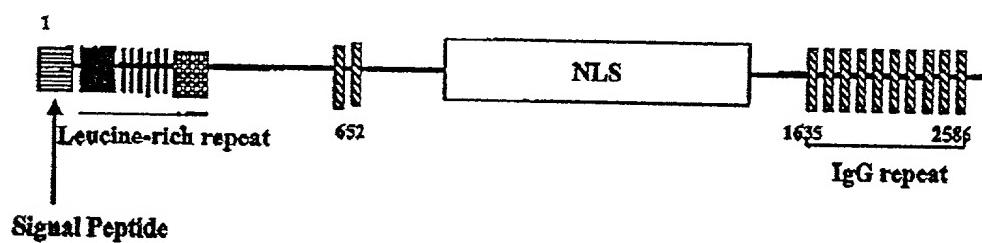
Figure 20
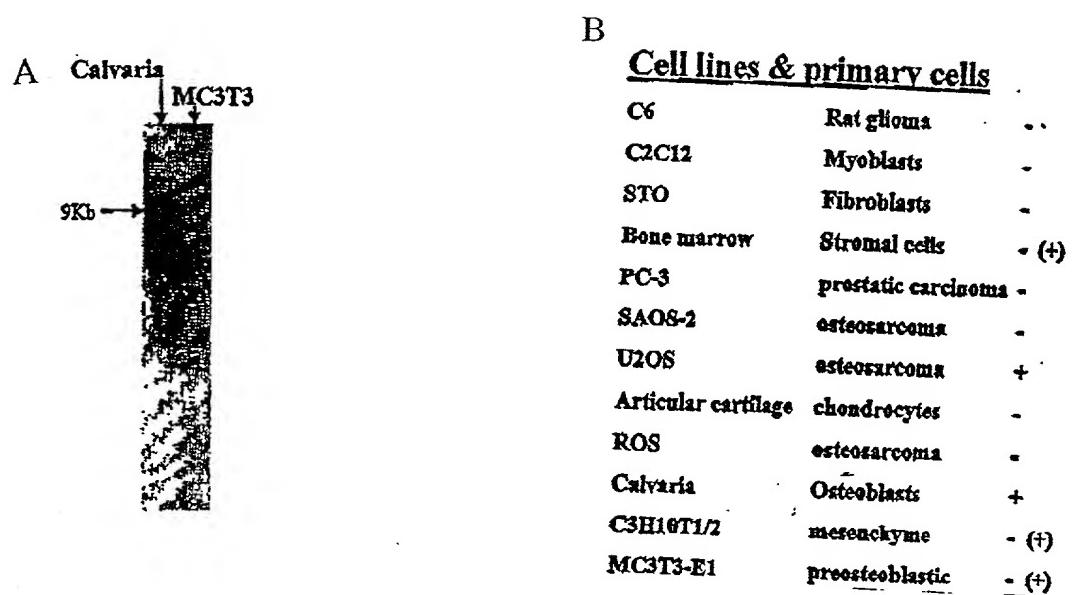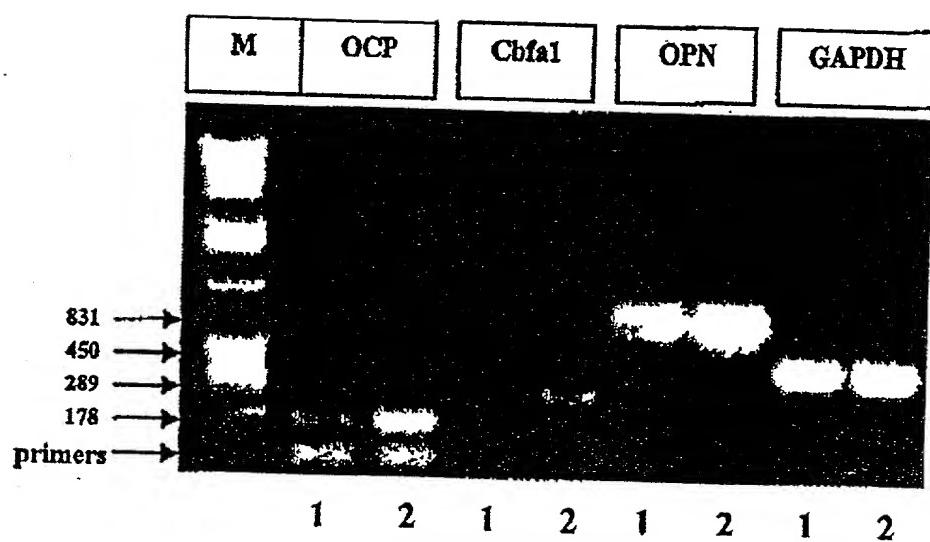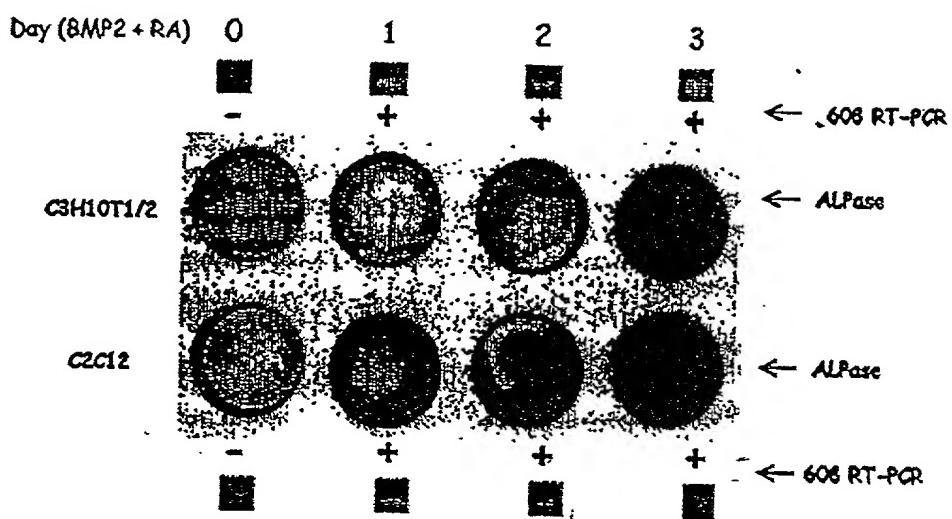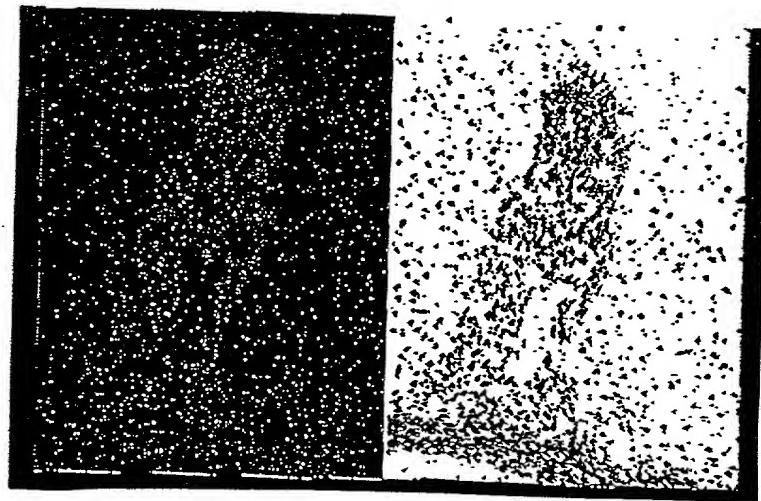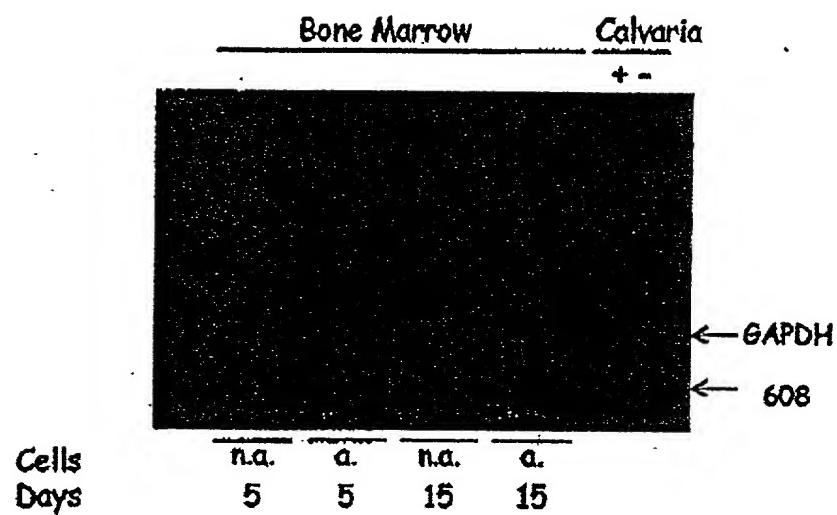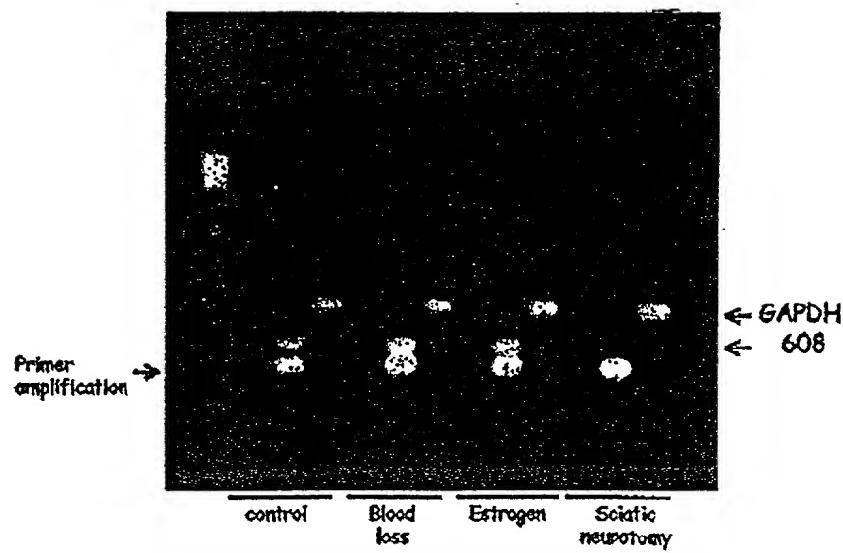
A
Calvaria
MC3T3

9Kb →

B
Cell lines & primary cells

C6	Rat glioma	· ·
C2C12	Myoblasts	-
STO	Fibroblasts	-
Bone marrow	Stromal cells	- (+)
PC-3	prostatic carcinoma	-
SAOS-2	osteosarcoma	-
U2OS	osteosarcoma	+ ·
Articular cartilage	chondrocytes	-
ROS	osteosarcoma	-
Calvaria	Osteoblasts	+
C3H10T1/2	mesenchyme	- (+)
MC3T3-E1	preosteoblastic	- (+)

Figure 21

M	OCP	Cbfa1	OPN	GAPDH

831 →
450 →
289 →
178 →
primers →

1   2   1   2   1   2   1   2

Figure 22

Day (BMP2 + RA)   0   1   2   3

← .60B RT-PCR

C3H10T1/2   ← ALPase

C2C12   ← ALPase

← 60B RT-PCR

Figure 23

Figure 24

Figure 25

Figure 26

Figure 27

A

B

Figure 28

A

B

Figure 29

A

B

Figure 30

Figure 31

Figure 32

A

B

Figure 33

Figure 34

Calvaria

Protein permeable membrane

Calvaria + 608/+pcDNA

pcDNA

608

Figure 35

| pCDNA | OCP-403 | OCP-760 | OCP-full length |

x10

x4

Figure 36

Figure 37

Figure 38

Bone

Calvaria + 608/+pcDNA

Figure 39

Figure 40

Figure 41

A

1020

B

450

A

Figure 42

	Calvaria length	Calvaria width
□N	89.06	56.83333
■T	96.6	63.25

B

	Calvaria length	Calvaria width
□N	100	100
■T	108.5	111.3

Figure 43

608

Control

2 mm

Figure 44

Control

Transgenic
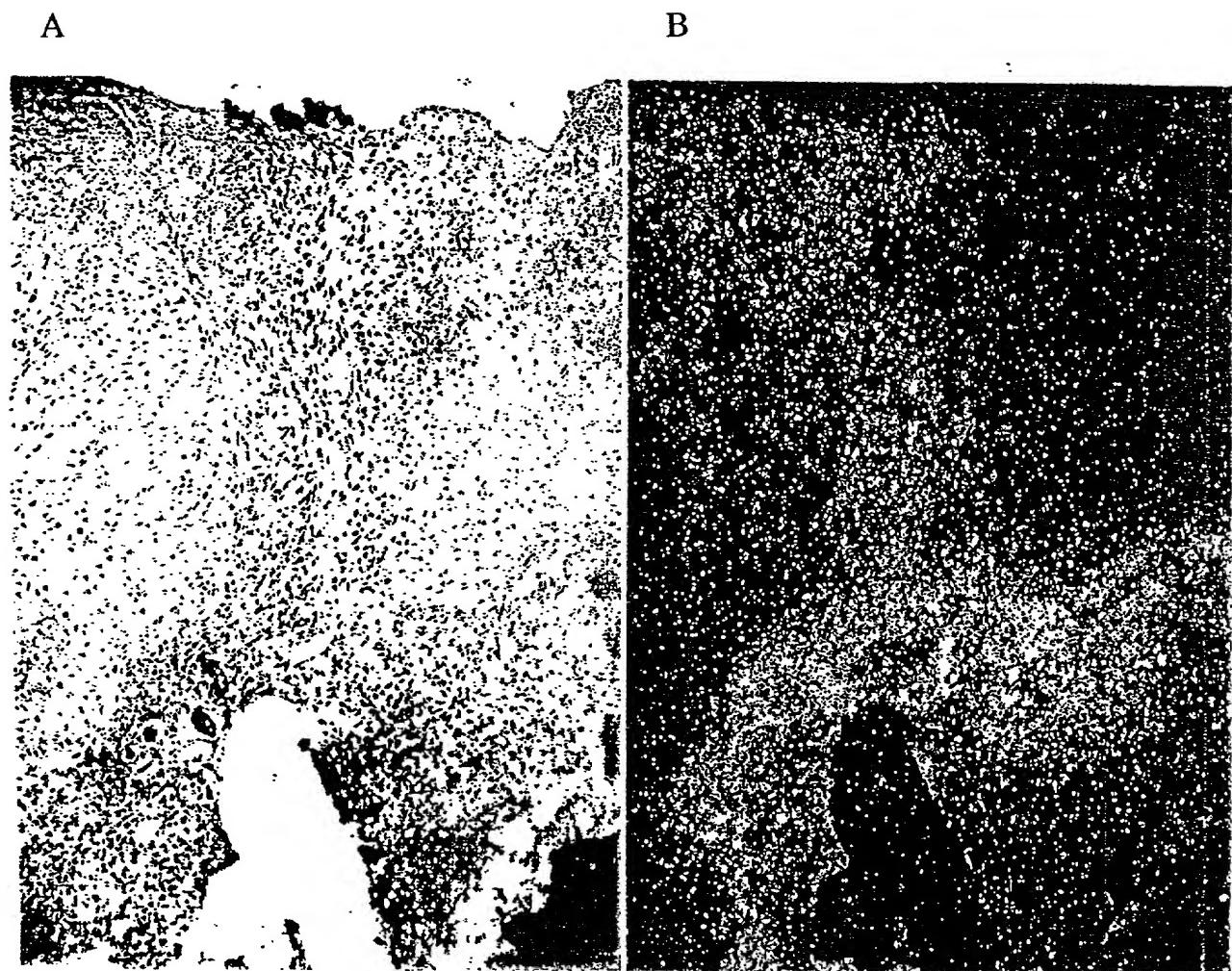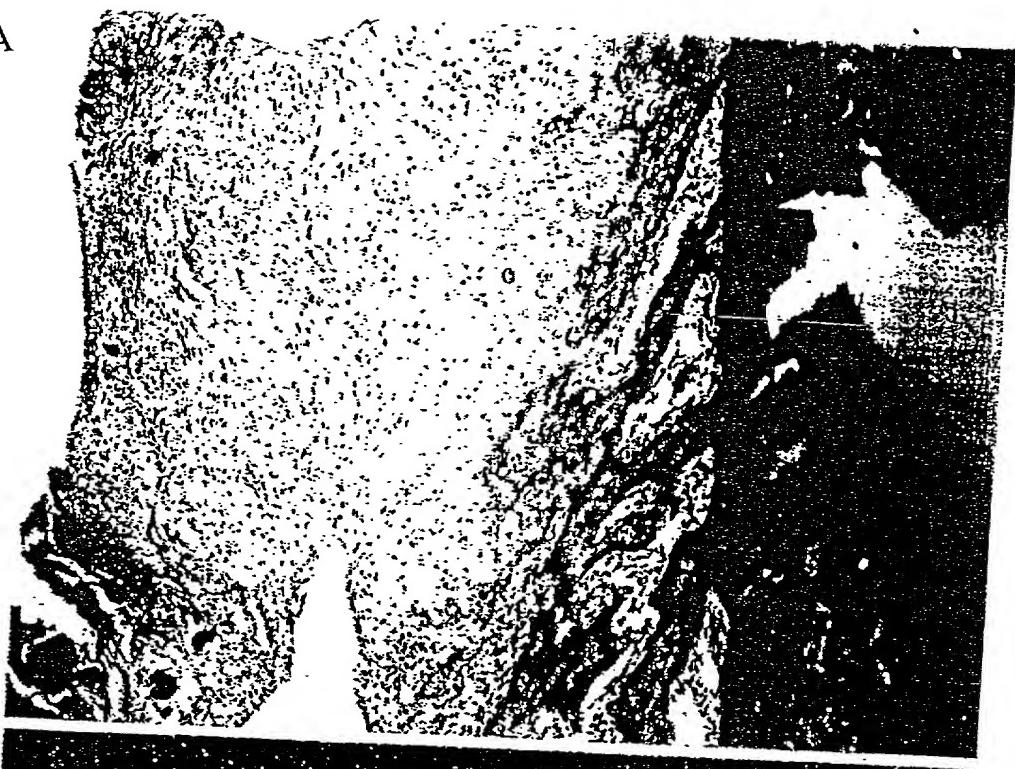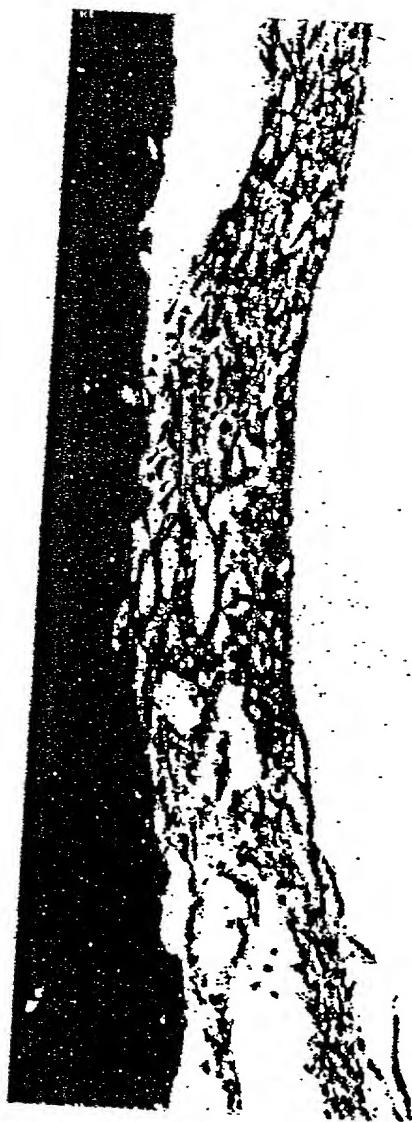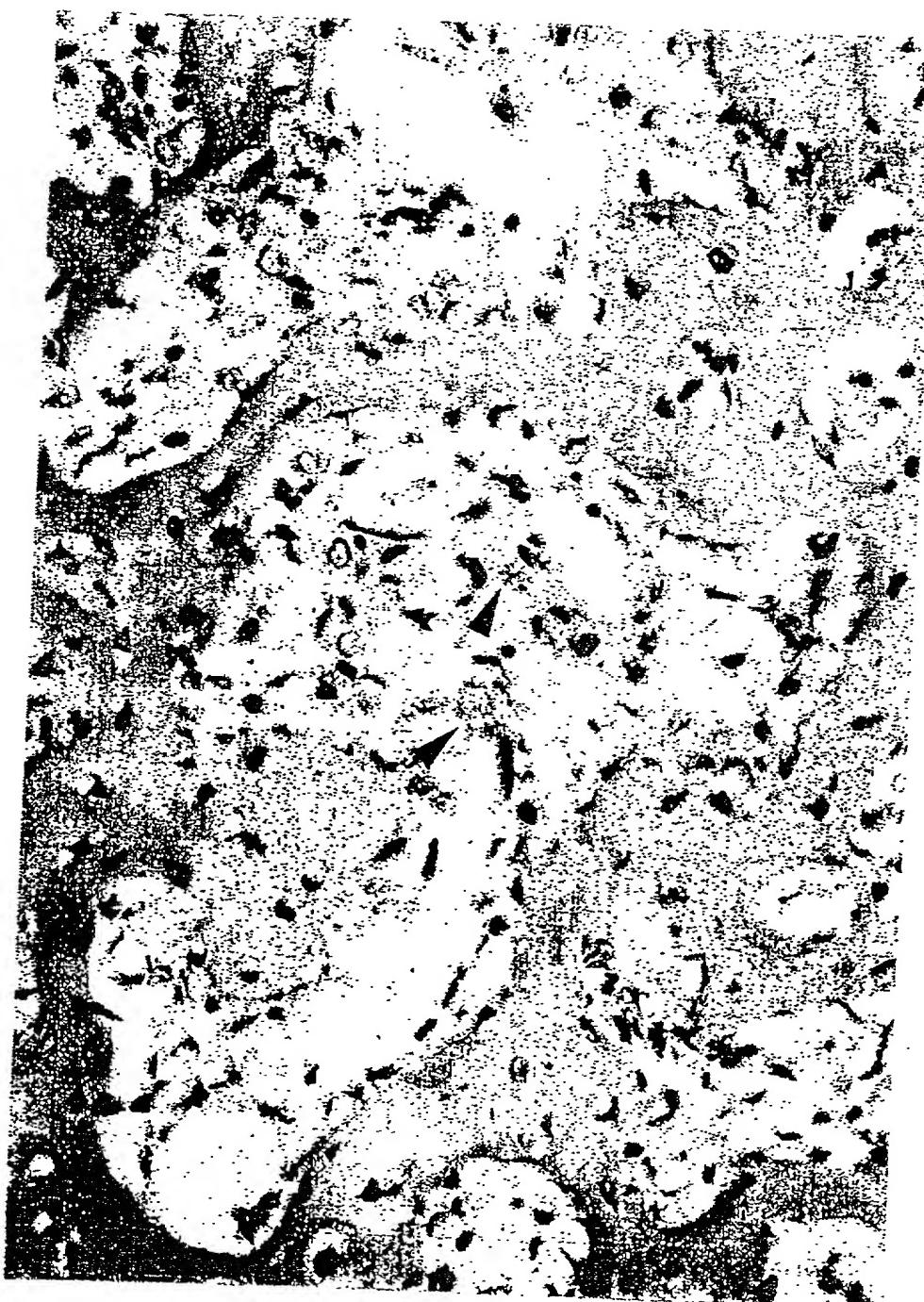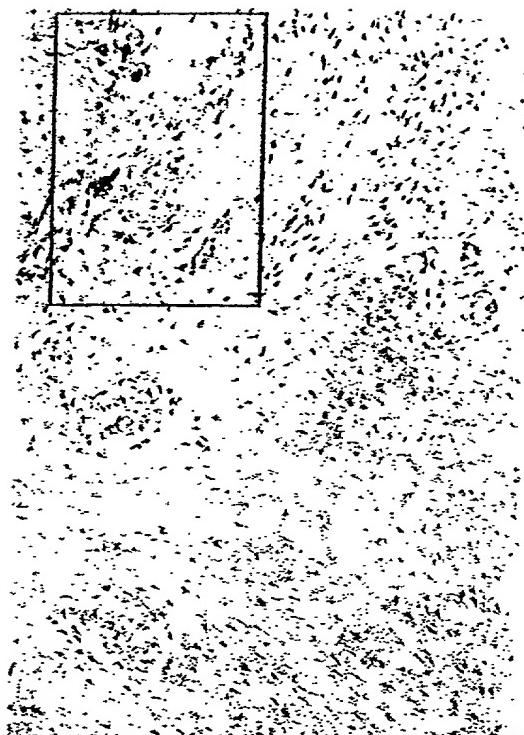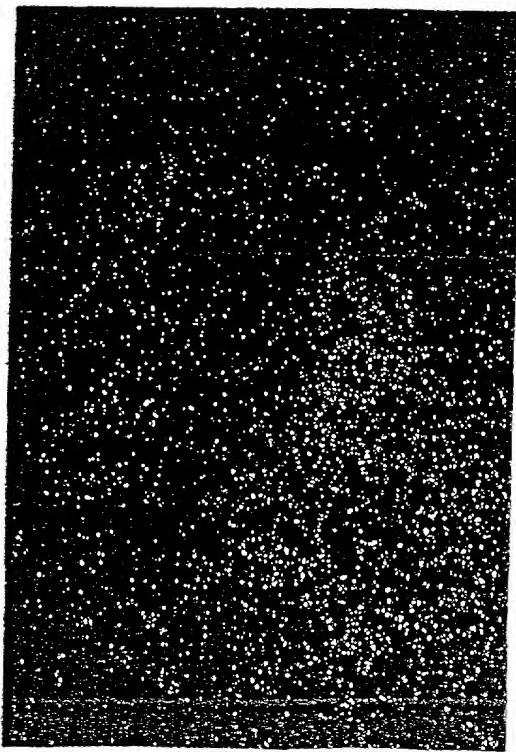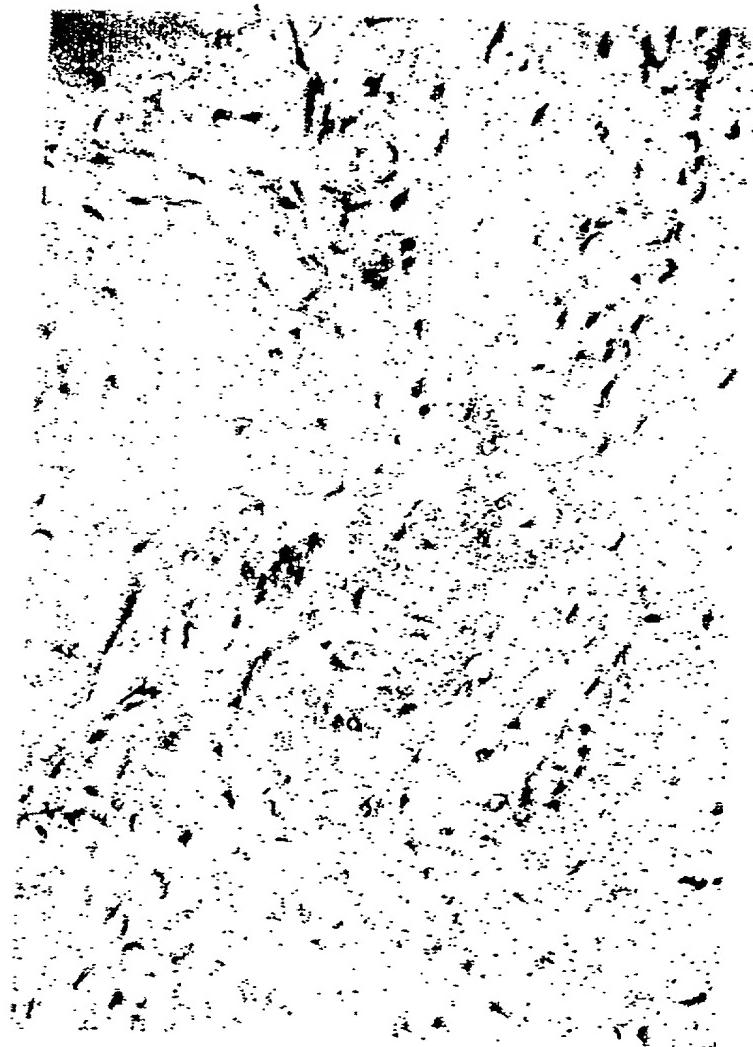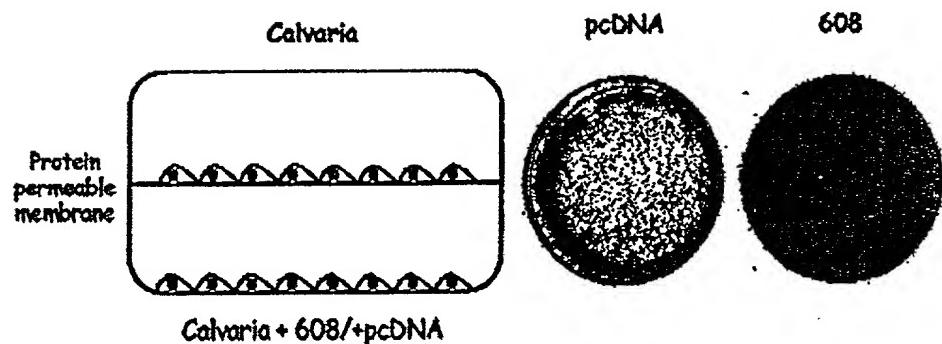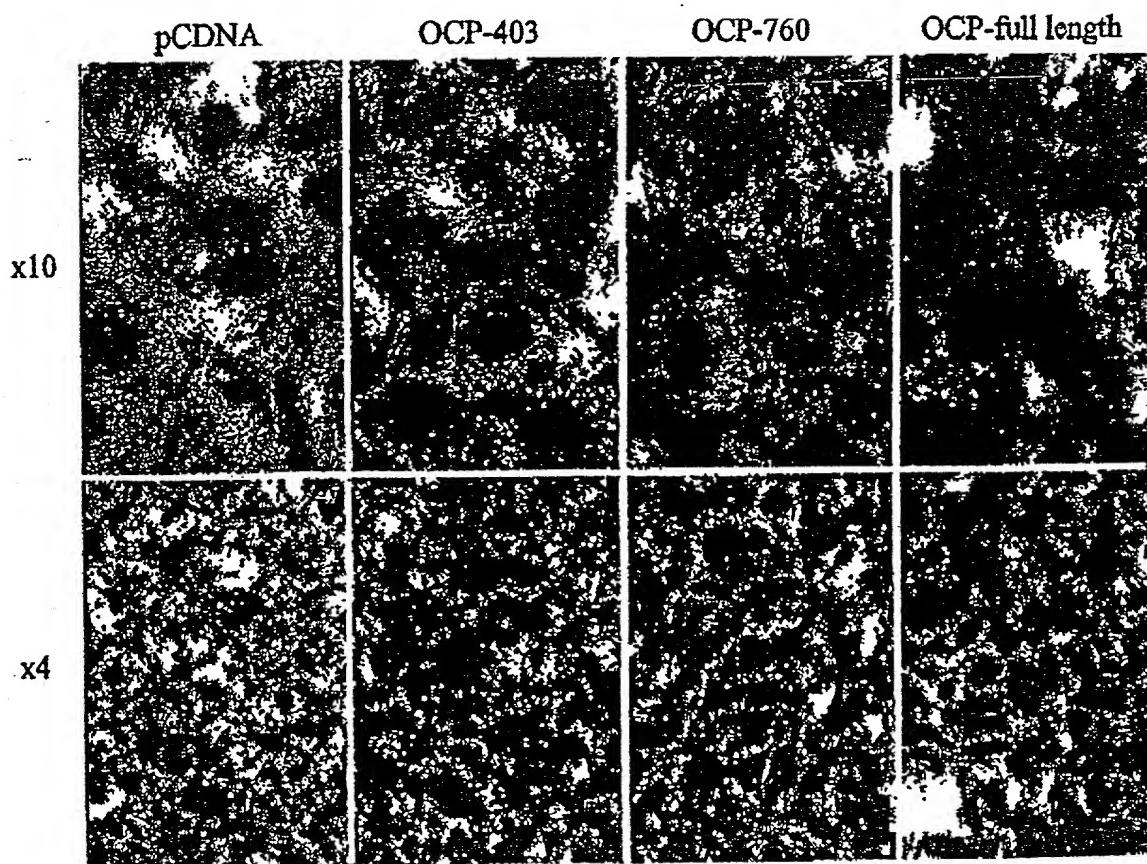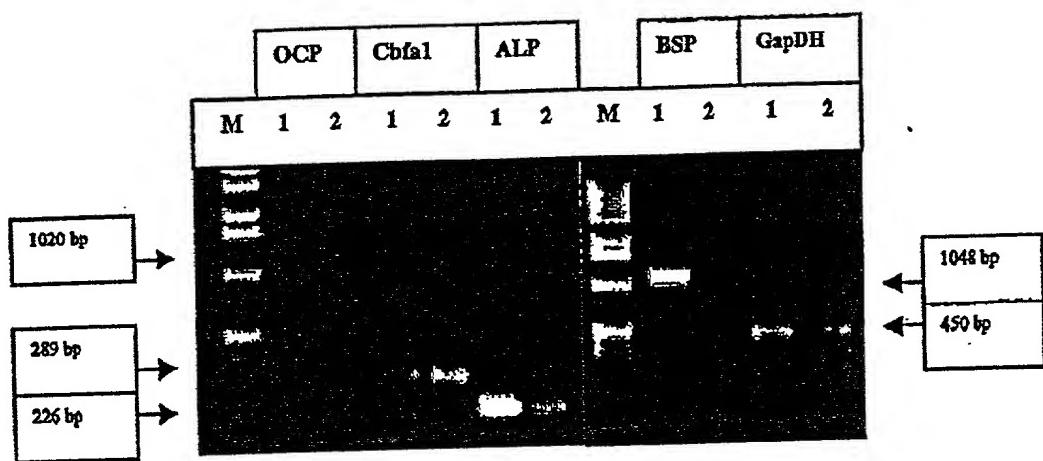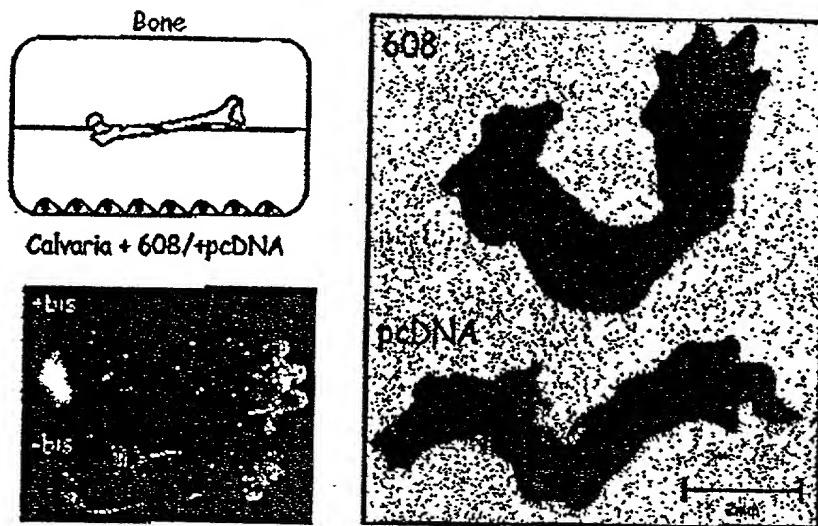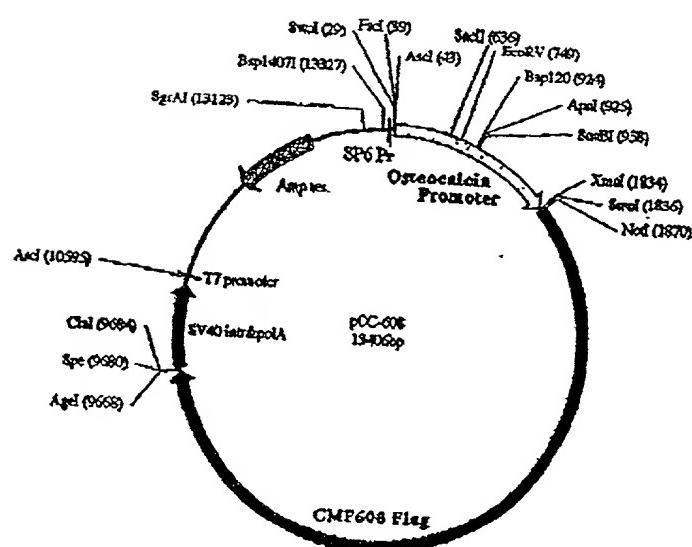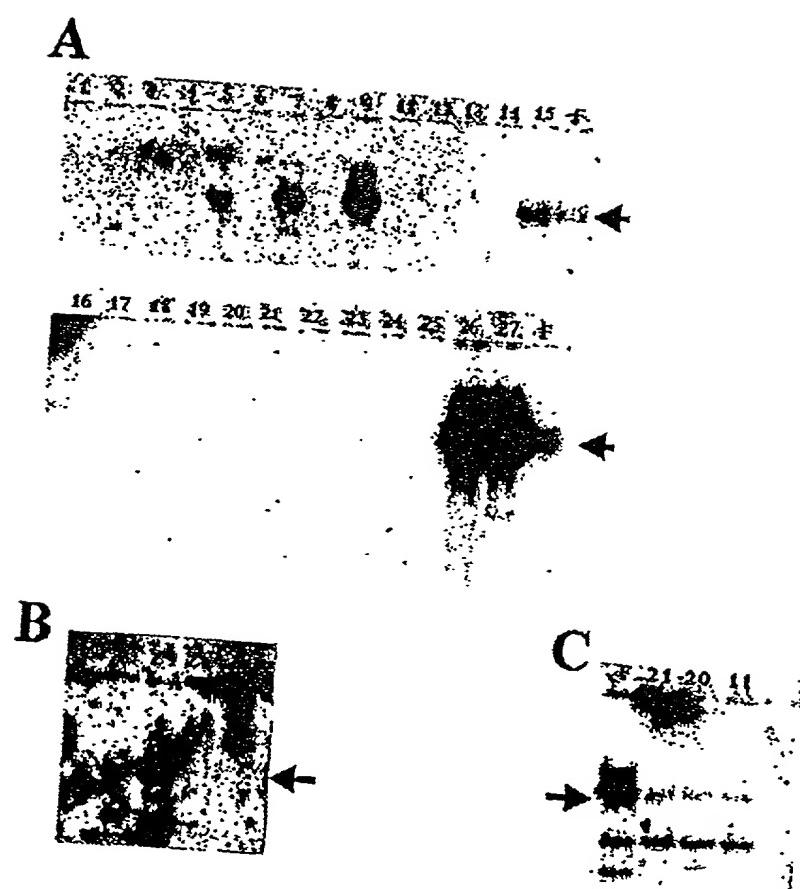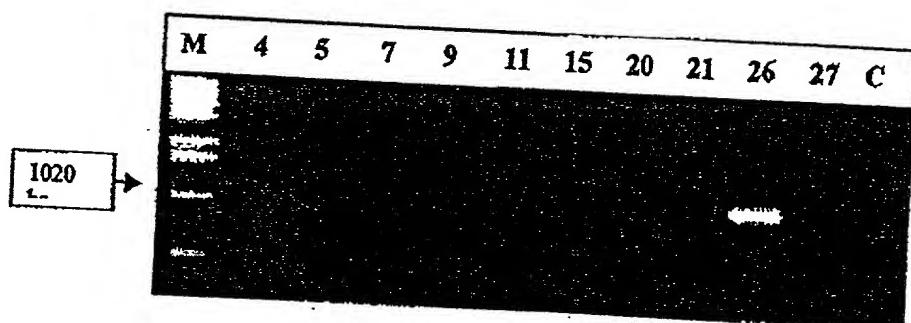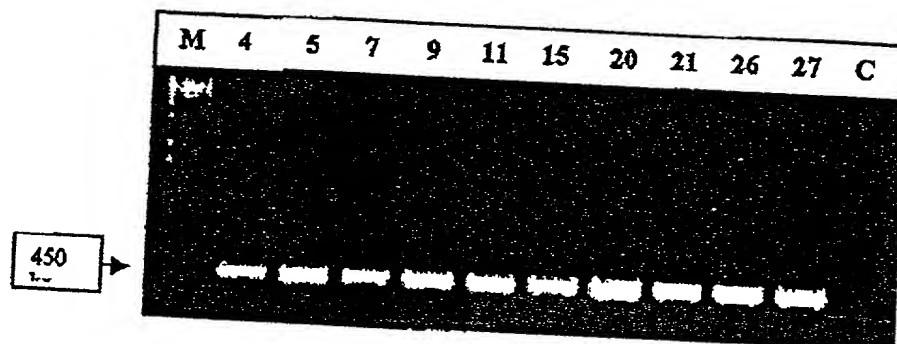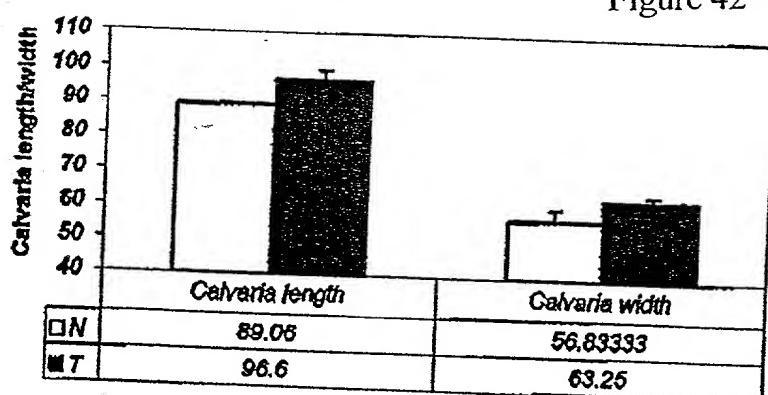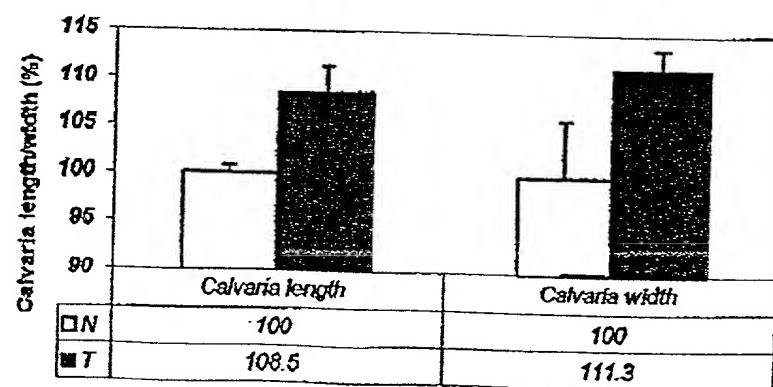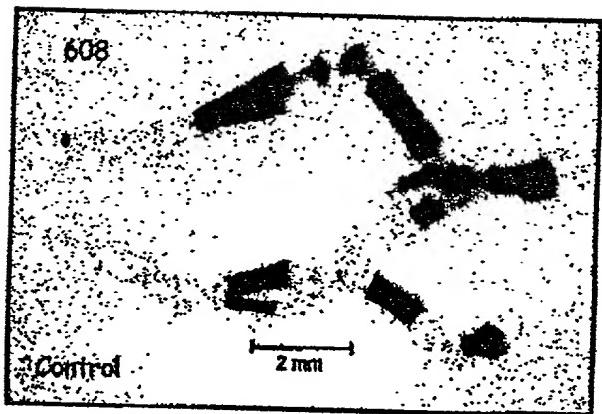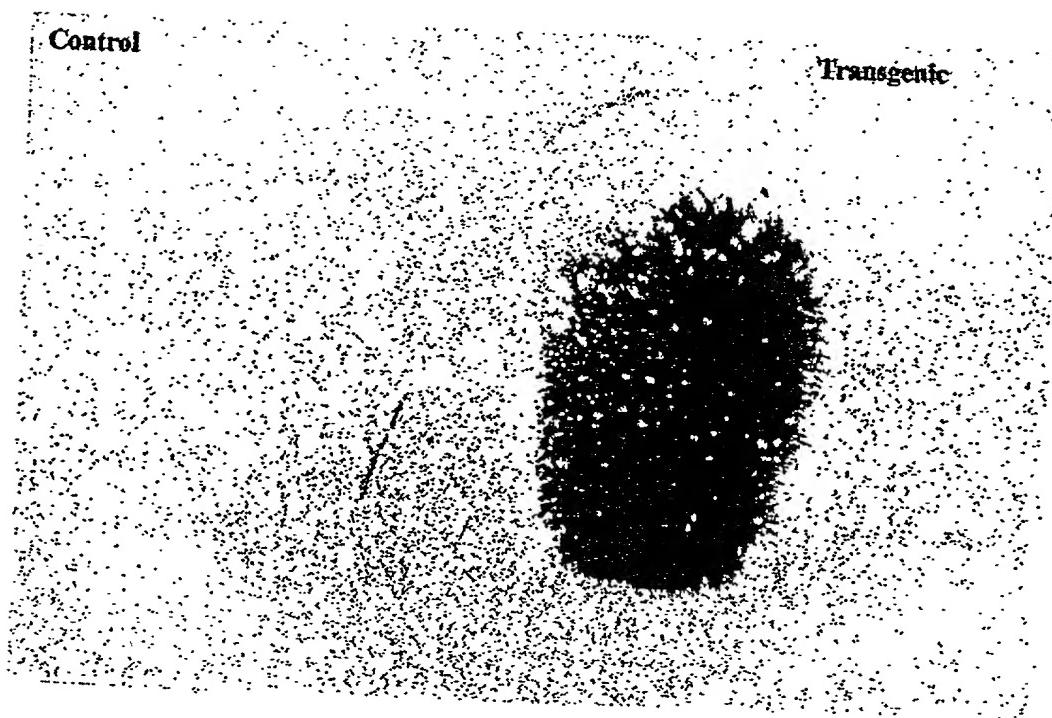
Figure 45

Clone 14C10 /15E11

Figure 46

Figure 47

↓ (XbaI)

TCTAGAAGTAAAATGATCCTGAGTAGCGATCCTGGGAAAATACGTACTCTAACACACTGC
AATCATCTCTCTGTGGTTTGCTGGAGCTGAGGTCTGGAAGGCTCGACCTTGGTTAGAAATA
ACCTACCGAATACAGAGCTATGACGTTAGTCTGGAAGGAGCTTTGGAAGAATGACAAGCT
GTAGCTGCCCAGAACATACTAGATGCCATATTTCCAAGGCAAGTGTCCACATGCGGACAT
CTTAAGAATATGGTTGTCTCTGCAGTGCTAAGGACCTTGTTCGTGCCACACAGGTCTCCAG
GGTTAGTGCTAACTCTGACTGCTTGACTCTTTAATTCTACCTTGATCATTAATGACTAGAA
ATCACTTGGTGATTAGCAACTGGATATGGAATATTACTAATTTGTACCCAAGCCAGGCCAC
CTCAGCTTTGGCAGCTCCATTCATTCTGTGGAGCCCAGTCACGTGGGTTTGAATCAACTGT
ACTGTTTCTACTTACAAGACGCATTACCTGAGATGAGTCATTTTTCTTCACAAGTCTTTTTA
GAAGAGTCAATTAGACATATTCTGATGAAGTAAGCATATAAAGTGAGAGCAGCATGAATG
TGTTCCATGTATGCTCATGGATGCTATTATAATGTGGAAATAAACTGACTTTAAAAAAAAA
AGCTTATGATACTTGTCACAGAGTAAATCTTCCATAAATATCATCTGCATTTATAAATTAT
TTTCATAATCCATCAATTAAAAAACCTTTAGAAATTTTGTTAACACAAAGATCCCTAGGCCC
CTGCCCTAGGATGGTCTGTATGGTGGGCCTGAGAGATGGAGCTTAAGAACTTACTTGCTCC
AGGAGCACATCTTCAGAACATCTGCCTCAAAACATTTATCCCAAATGCTCATCAAAGGCTC
ACTCACATGTGCTTCAACCACAGGGATTAAACAGTCATTTTAGTCACATTTCTCAAACGGT
GGAAGCCTGCTAGAGGAACAGGATGTATCAGGATAACATCCAACCTTACAAAAGGATGTC
ATAACCCTCACCACAACAAACAACAACGACAACAAACCCATAAAAATTATCACGGCAAAT
GAACTAAGCCATATGCAGAAAAAGTATTATATGTTCTCATTGTGGGGTGTTTTTCCTTAAT
AGTCAAATATGCAGAATATAGACAAAGATGGTTTATGCAAGTGGGGATGGCGAAGGATA
CTTGTAGATTAGAGGACACAAAGCAACAACTACAGAGTGAAGTAATCCAGAGACTTAATG
TATAATATGAGGACTGTATTTAATAATTCTATTTAAGATACACAGCAAACGAGTGTATCTT
ACTAACACACACACTTACATAGAGAGAATAAAGTGATAGATACGTTTGTTTTATCTTCATG
TAGCTGATAATTTCATATTGTACACCTCAAACATAGATAACCAACAAAGAGGAAGAGGAT
AGGTGCCTCTCCCAGGGCGGAAGAGTACATTCGAAAGTCAGACACCATTGTGTAGATGTA
CCACATGGAGGAGCTAGAGAAAGTAGCCAAGGAGCTAAAGGGATCTGCAACCCTATAGG
TGGAACAACATTATGAGCTAACCAGTACCCCGGAGCTCTTGACTCTAGCTGCATATATATC
AAAAGATGGCCTAATCGGCCATCACTGGAAAGAGAGGCCCATTGGACTTGCAAACTTTAT
ATGCCCCAGTACAGGGGAATACCAGGGCCAAAAAGGGGGAGTGGGTGGGCAGGGGAGTG
GGGGTGGGTGGATATGGGGGACTTTTGGTATAGCATTGGAAATGTAAATGAGTTAAATAC
CTAATAAAAAAATGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAAGGTCAGACACCTCAC
TTCACTGCTATCTCAACTTGCAAAACAGAAGGGGAGTCACAAAACCCAGGACAAACCACAGT
GATTGAAGCGTCTTTGAATGTTATTGCTGTTGTTGTTACCACCATCATTAGCATATATTCAT
TGTGAAAACTTACGGGGTCTATGACATGTTTTTTTATTCAAGTATATCACATGCTGTCAGC
ATATTTGGCACCACTACCAGCCCCAGCCCCCTTTGCCCCGCCCCCAACACACACACACACA
CACACACACACACACACACACACACACACACACACACACACCTTTACCTTCTCCTGGG
CATCATCTGCTCACTCACCCACCCAAGCTTAATCCTTTTCCTTCCCTGCAATAGTACCTCTC
CTATTTTTATGTCTAGGTTCCCCCTCCCCCTGTTAGGAGATGGGAGAGGTCACGAAAGGAA
AGAATTTGTAGCCCCTGAGCCAGCCCGGGCCACAGAGCCTGCCACCAGACAGGAAAAGCC
CAGGGCTTACCAGCACAGGAGGAGCAAACTCGCAGGCGAGCCTGGGTTGGCGCTGGTGGT
CCCGGGTCGATGGCCCGCCCATTCCCAGAAGCCGAGGCTATAGCTGCGTCACCTGCCCCG
CCCTCCTCCCGAGTGAAGACCCCTAGAGGCTGAGCAGACCCCAAAGGCGGTGCAATTCCA
TTGGCCCAAGGCAGAGGTGAGCGGCTGCTAATCCCCTCGGGAAGTGAAGGGACCCAGAG

Exon1
AGTCTGGTAGATGTGGGAGCTGGGGTTCAGGGCGAGACAGAGGGTGGGATGGGCAGAAG
GGTCCAGGAAAAGGAAAGTACTGGAGGGGAGTTGGGACAAAAGCAGCGACCAAGGGAA
CATCGCTTCAGTGACTGAAGCCAGGCAAAAGGAGCGGGAAGGATTATATGTAGCCTGGGA
CGCTTTCATAAACACTGATGACGTGTTTGTGCAAAGCAAGCAATTTGAGGAGAAACGCCT
GGGACGTCGGAAAGAAGGAGTGATCGATTAGTACTTGTAAGTTTAGGTGAGTTTGAGAAC
TAACTAACCTATACTATTGAGGGAGAAGGAAGAGCATTCCAGCAGCAGCAGCAGCAGCA
GCAATCAGATAAAGGAAAGCTTTGTTAGTTTGGAAATGTATGATACCATTAAAATAACA
GAAGCGCCTCCAGTTCTCTGAAGAGTCAGTCCCCCAGCTAGTGAAGACTAAGCCTACTAA
GCCTTTTGCTCCCGTTGGAAGCAAAGAACGTTCCTTCAATCAGGTGAAGGCTCTCCTCAGA

<u>AGATTTCCTGTCTCTGCTTATGTTACAAGAGGATTCAAAAGCAAGACAGAAGAGCTCAGG</u>
TATTGCCAACTCTTTTGTTAAATACAGTTTGAGGCTTAAGTGTACGGGAACTCATGTGGTA
TTCATTTACGGCTCTCTTCTCTTATAACTAACTCTTAAGGTGCATATAGTCTCTTCTGTTTCC
CAGCTACCTTGTACCATCTTTGTTTATCTAATAATAGCAAGCTCATCTGCTTTTTAATCATC
ACGCAGAGAGTATTCAAAAATATTCAGTGATGTAACAGTGACAGTGTAGGCATAGAAGTA
ATCATTAGTAAATCTTAATTTGGGTTAAACTCATTCATAACAGCTCCAGGTTGGGAGGGAT
CACTGAGCCTTCGCCACGTGCGGGTTAAAGATATTTTCTAACAAGAGAAGCAGAATTCTTC
CTTGGCCATGCTCCCCATCACTGTGTCAGTAAGCAGAGGGGTGTTTCCAAGCAGAGAAAG
AGCAGACAGTGTTATGCCTGCAAAGTCAGAGACTCAGCCCTCCCAGCTGGTCAGTTTACT
GTCCTCCCGGTCATTAGTTGGCTCTGAAAAGGCCCATGTGTCCTTATTGGCAAGGACTTGC
AGACATGCTAGAAAGAAATTTGACCTTTTTTTCTAGTGGGTTATTACAGCTGTAAAAGTAT
TTTGGAAGGTTAAGCCAAATAAATAAAACACATATTAAATAATACAATGTTACAAAAATT
GATCATATAAAGAAGTACATTCATAAATGCAATGTGAAAAATATATATAATTTTTATCTAT
TTACTGGTGCAAAGTTTTCTAAATTGCACATGTACTATTTTTATATTTATAAAAATATTTTT
AAAATGTATATAAAGTGTAAAAGGCTCTTGGTCAAACAAGAGAGTTAAATTTACAAACT

                                      ↓ (BamHI)
TTAATTGTCCCGATAACATTATTATGATCTCTAATGACAG<u>GGATCC</u>TGCTTTTCATTGGGA
AATGAGAAGCTATGAAGATATGTTTACAATAATAAGCCCATTTAGTGATAAAGTCCAATG
GGAAGCTAGCACACACTGGTTTATAAAGAGAACAGTTTCCTGAGTCTATGCAAGTTTACA
CTCTAGGGAATAAGAGTTCCTCTTTCTCCAGATTTCACTAGCATTTGTTGTCATCATTTATC
TTCTTGATGATGAGCATTATAAGTGGAATAAGATAGGATCTCAAAGGAATGTCAATTTGG
ATGCCCTGAACAATCTTTCAGGTCTTTCTTTCAGTTCACTAGTCTATTCATTTATTGGATAA
TTGGGGGATGGTGTTAATTTTTTTGCAGTTCTTATGGAATTCCAAAAAAACAAAAAACAAAC
AAACAAACAAAAAACCTCTGAAACTAGAACTACCAATCCATTACTGGGTATGTAACAAAG
AGAAATCTGCACAGAATTTATTGCTACATTGTTCATTATTCACGACAGCCAAGAATGTGGA
ACCAACTTACGTAGCCGTCAAAATATGAACGGATAAAGAAAATGTGGAAATGTGTACAAC
AGAGTCCCATGTGTGGCCATAAAAGAGTGAAATCATGACATATGCAGGAAATGGATGCAACT
GGAAATCAATTGGGCTAATCAAAACAAGACAGACTCAAAAAGGAAACACCGTGTAGCTT
CTCTGACAAACAGAAGCTAGATTTACACTTGTACGTGCGCATGTGTGTTTAGAATTTTATT
TAGTTATACACTATTCTAATCTGTGAGTGTGTATAAAGGCATGCATGTAAAGCAAAAACA
AGCTAGCTGGGGTGGGTAGGAGAGAAAGCAATGAGAGGGAGTTAATAAGAACGAAGCATA
GTAACATAGGTGCCAGGATGAAATGCATTAATTTGTATGCTAACTAAACCACAGACAGGA
GGCACACGTTCAAACCAGGGTGAAATCCCAGCACAGAGAAGGGGAAGTAGACACAAAGT
TTCGCCACTAACCAAGAAGCCATTTGCAGTTGCTGCCTGCTGGGAGGGGCGTTCCAGTTTT
CTCCAGTCTGACACTGTGTATAACAACCAGTTGACAATACAAAGTTGGCATGATGGATGG
TTTTTGTGCTATTTTTCATTTTTTTTCTTACTGTTTTGTTGTTGTGGTGGTTGTTGTGGTGGTG
GCTGTGGTTTTCATTTGTTTCTTTTGAGAGAGAGAAGGAACATGAAATTGGGTGGGTAGGA
AGCTGGAAACGATCTGGAAGAAGTTGGGGAAAGAGAAAAATTGTATGGAGCATATTTAA
ACAAACAAACAAACAAACAAAAGGTTCATTTGCCACAAAAAGGTGTGAATTAAATTAAC
CAGTTACGACTCTTAAAGAAAATATTCCCAATTATTCCCAGAGTTGCTATGTATGCTGTGC
CTAGGACTTTGCTTGAACTGGCCCTATAACTCTGGTGTGGTGTCTTTTCAGG – **Artificial**
                                                   NotI site

(SEQ ID NO:17)

Figure 48

CACAGACCTTCCTCTTCTAACCTCTCTCCCCCATCTTGTTGCTTCATCCCAGACTTCAACAC
CAGCAAGCACACTCTGCTAATGCAAGGGCTGCTCCTGTCAGGACAACAAGGAGGCTGAAG
GCAGACCCACACGTTTCCAACTGCTCCTGAGAGTCAATCCCCCTAGACTCATCTATAGCAG
GAAACCTGCTGTGATCTCCATTTCTTCTCTGACCACATCCCCAAGTTATCACAAGGAGTTT
TTCCTCAAACCTTTCCTCTCCAGCAAACCCCTTCAGCTCCTTGGGTACTTTCTCTAGCCCCT
TCATTGGGAACCCTGTGCTCCATCCAATGGATGGCTGTGAGCATCCACTTCTGTATAGAAT
CTTGGTCAGTGCAGTCTTTTGTATCCTCAAGAACACTGGGTCTGAAAATTTTAACCCAAAG
AACTGTTTTTGTTATGATTGCTGCAATCTCTTTCAATTCCAATAAAGAGTAAGCATCTCAT
TCCTTTGTCTCCTCCTTTCAGTACCACCCTGCCTTTGCTGCCTTTCTCAAAGAATCAATAAA
ACCAAAGTGATATAGATTCATGGCATTCCTCTAACTGCTACATCCACTCCAGTAGTATCTC
ACTTGGCAGGTGTAAAAGCCTGGAAGCAGTCACGAGGCAGTTTCACAGAAACTTAGCCTC
CTGGAACCTTGGCATTCCCATAGCTAGAATGCCCCAGATTTGTCCCTGAGATATTGTGGTG
GGTCTTGCATGCTTTCTTGCAGTATTTTACTGGATAAGAGTTAGAAATCTCAGGGCGAGCT
TAGCAAAAGTATACCTAGAATCTTCATGACAGTCAGGTATTGCAAACTACATTGCATATTA
GAAGAAAGTTGGTAAATTCTTCTGACAAATGGAGATTCCCTACAGATAACTTAAAAGAAC
AGCTAAGTCACACTCATATGCAAGAATTTACCAAGGCCTAGGAAAGGGGGGGGGGGTACT
GCTTTATTCATGATAAGGTCTGCTAGAGCAGAACCCCCTGGTGCTAGCTTTCACAAGGTTC
AAAGGTGTAGCATAAATTGTGACTAGAGTGTGAAATCTTTACCTGTCATTAGCTGACTCTA
GGCAGAGCTGTTTTATCTTTACTGTAAACATTACCTGGTTCCTGTCAGTCCTTTGAAGGCAT
TCCTCTGTTTTGTGACAGATACTTCTATGTACCTCGCCTGCTGTGACACCCTACTCCTTTGT
TTTCTGTATTATATAAGCCTGGTGTTCCCTTTGTGAAAAATTACATCCAGATACAGCACTC
CCTTGTGTCTGTGTCCTTTTGTCATTTCTGGCCAACTCCATGCCCACCTGCCAGAACCCCTA
GTCTTTTCCACAGATTGAGGGAGGCCGACTGAGCCTGGTCCATGGCATCTAACCACTGTCA
GCTCACTGTTGGTGACTACCTCAAGGTACAAGCTCCATTACTAATGAAACAAAATTAGAT
AAGTGTGGGTCCAGGAAGCAGGTTGTACACCCTGTCTGAATGAACATTATGAAATGACTG
AAATAAGTTAACCCATCTCTTCCTCGTTTGCTAATATAGCAAATAAACCGAGTTTCTGAGC
TGCTGCTGGTGTGTCTCCATCAGAGGGCAGAGCCAGTCTGATCCTAGCTTTCCTGTATGTG
TGTCCATTGTTTCTTCAGTTCCTGTTGCCCCATTAGGAAATCCTAAGCCATGAAAGCCATG
AATCTGGGAATGACTTTTCTAAGAAATGCCACGTGAACCTTGCGTTTCAACGTTTTGCCTG
TAAACAAGATATATGGTGCGCAGTTTATAATCATAATAAGCTTTGAAATAATATATAACTC
CATTCTCATTCTGCTTCCACGCTGAGCATCCTGTTTCCCCAGGGACCACAAGAGCATTTGA
AAAGTAGTGATTTATGACCTGCTTTGTTCTGTTACTATAAAAGCTTCATGAAAGGGCAGCC
ATGTTGAAACATGGAACTTGGGGGTGACCTGTATCTGTGTTCCTGGGTCGTGCTCACTCATA
TTTGTCTCCAGAATAAATGAGTTTATCAACTTCGAGGAAAAAGTTGTGTGTTTGTATAGCA
CGCCCGTGGAGTCCCACCCATTCTACTTCCTGTAATCTGTATATGGTAGAAAAAGTTAATTT
ATGTGATTCTTCCAACTCCAAATATTTCAAATCTTTTAGCCCCTCAGCCTGGGATTTCTTTG
ACTAAGTCTATTGATTTGGAAGATCTCAGTGGTTAGGATTTGCAGTCATGATGTTCATACG
TCAGGCTAAGCTGAAAAATATGACAAATGAAATGTCAAATGTCATGTGCCTGGGAATGTG
AGTGTTAGGGGGTTTTAAAGAAACAAATACCTACTCTAAATAGTTAATAAGTCCCATGGTT
CTATTCTAGTTTTGAATAATGTTCCCTAGTATACAGCAATTTAATTTGAAATGAATAGCTTC
TTATCTTGACCAATCTCAGTGACTTCATCCGTCCCAAGTCATGTTTTCATATTCATAAGGAT
AGGTCTCATTCAACCACATGTTTATCATTTGGGATCTGCATTTTTCTGATGCAAAATGATTT
ATTCTTCCAGAGCACTGGAATTGGGTTGAATCATCTTATAACGGCCAAAACTAAATGCTTC
TGTGCTAAACAGAGAGTTACAAGACCTTTTTATGTGGATGGCAGCATTTTAGTCATCCTTA
TGACAGAATGTCAGAGTGGAGCTCCCACTGGGGGAGGGGCTGGTCCTTGGCAGGATTCTC
TTGGAACATCACACAAAGAAATTCCAAATTATGAAATGCACATGATCCATCCAGAATGTG
ACTTTTGACTCTTGAACATGAGCTTTTAAAGTACGTTTGGCTGTTCAGACCTTGACTTTGAG
GTGAAGGAAAGCTCGCCAACTCCTTTTTATATGTAACACAATATATCAAGATCTAATGTGA
GACAGTATGCCAGTCCCAAGATCTGTCAATATGACTGAAGACACATTGCGATGTTATCACT
AAGGCAGGAGAAGGCAAGCTACAGTGAAGCCCAGTTCACTATAAAGCTTTATGAGAAATT
AGATAAGAAGGGTTTCTAATTTTTAAATTTTTTTTTATTAGATATTTTCTTCATTTACATTTC
AAATGCTATCCCAAAAGTCCCCTATACTCCCCCCCTCACCCTGCTCCCCTACCCACCCACTC
CCATTTCTTGGCCCTGGTTACTTGTGATAGTGGTCATATGATCCACCAAGCTTTACATGCTC
ACTATCTGGTCTATTGCAAGAATGGCTGCCGAGCTGATGCAGTCAGATACAGACACCTAC
AGCCAAACAGTGGAAGGAACTTGGGGACTCTTATGGAAGAAAAGGAGGAAGGGTTATGG

GCCCCGGATGGGGAAAGGAACTCCACAGGAAGACCAACATACTTGGTCAACTAACCTGG
ACCCTTGGGGCTCTCAGAGTCTGAACCACCAACCATAGAACATTCATGGGCTGTACCCAG
GCCTCTCCACTCATATGTAACAGATATGTGGCTTGGCCTTCATCTGGGTCCTGAACAACTA
GATGGGGGGTTAGGGGTGGGGATGGGGGTTATCTCAAAAGCTGTTGCCTGTATGTGGGATA
TGTTCTTACTGAGCTGCCTAGTCTGGCCTCAGTGGGAGAGGAAGCACCTAGCTTTAGCCTT
GTAAAGACTTGAAGTTCTGAGGTGTCGGTGGAGGGTATACTCAGGGAGGCCCTCACCTGC
TAGGAAGAGAAGAGGAGGGGGGAAGACTTGGGGGAGGGGGCAGTGAGCAGGTTGGTAAA
ATGAATAAGAAAAAAAAAATAAAAATAATTAAAAAAAAAAAAAGAATGGCTGCTGAGCCC
TACTCTAAAACCATTGCATCCCCCCCCCCCAATCATTCAGTGACTACGAATTAAAATCATT
GATACTAACAATAGATGTAGGAAACTATTGTTAACTTCTTTGTGACCACGAGTGGTATTTG
GAACCTTTTTTATTGAAGCTTTCACACAGAGCCTTGTTCTTTCATTTCCCTGTACATGCATG
TAGCTTAATGATGTTCAGTGAATTAAAAATAACAATGAAGAATAAAGACAACTGTATTTT
AAGGATTCTTCGTATATATTTAAAAATCTAAGGTGGTCACCTGGAAGAAATGTCTTCAGTT
TTTCTATATATGTTACTCTATCGTATGTTAATTAATTATATGCAATAATTCATAAAATCTA
CAACATAGTATGTAACTTATAAGAAAGTAAAACATTCATGAAATTGTGAAGGTTACTTTTC
CTTACCCTCAGAAACACTGGGTTTGAATAATTCTTATTTTGGTATCAGTGAAGAATTTGAA
AGAATGTAATAACCTACTAAGGCAAACATAGAAGTTGAAATTAAAAAGAGTAGACAGGA
GAAGTAATAAGGCAAATAATGAATATTTGCTTTAAATAGTTCTTAATGTATCATCTAACTA
GGGTGTGATTCTCCAGACTTGACTCCATCCAAAATATCCAAAATGACTCTAACCACAGTCA
TTGAAACAATGTGTTGAAAATAATAAACATTTCCTACTTGAAAATTCAGATTTCTCCTACT
TTGCTTTTTATTGCTGTGATAAGCACCATGACCAAAGCAGCTTAT

(SEQ ID NO:18)

Figure 49

Human

608

ATG

Mouse

608

ATG

Cbfa1
MyoD
SP1

MyoF
WT1
SOX-9

Egr-3
Egr-2

VDR
PPAR
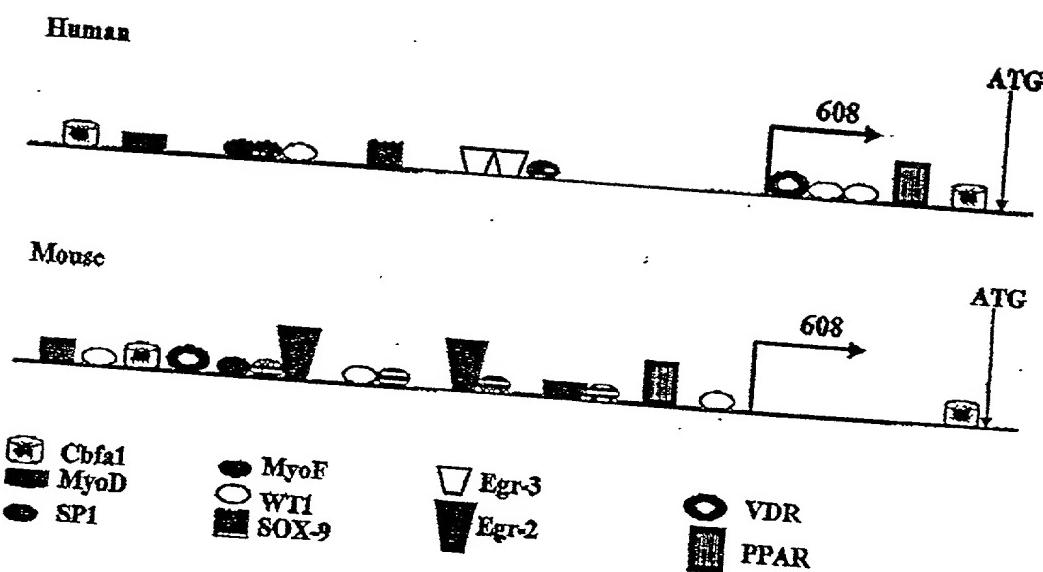
Figure 50

TAAGCCTTTTGCTCCCGTTGGAAGCAAAGAACGTTCCTTCAATCAGGTGAAGGCTCT
CCTCAGAAGATTTCATGTCTCAGCTTATGTTACAAGAGGATTCAAAAGCAAGACAGA
AGAGCTCAGGTATAGCCAACTCTTTTGTTAAATACAGTATGAGGCTTAAGTGTACGG
CAACTCATGTGGTATTCATTTGCGGCTCTCTTCTCTTATAACTAACTCTTAAGGTGCA
TATAGTCTCTTCTGTTTCCCAGCTACCTTGCACCATCTTTGTTTATCTAATAATAGCA
AGCTCATCTGCTTTTAATCATCACGCAGAGAGTATTCAAAAATATTCAGTGATGTA
ACAGTGACAGTGTAGGCATAGAAGTAATCATTAGTAAATCTTAATATGGGTTAAACT
CATTCATAACAGCTCCAGGTTGG (SEQ ID NO:19)

TTTGGAACCAACCCAGATGCCCCTCAACAGAGAAATGGGCCAGAAAATGTGGTCCA
TTTATCCAATGGAATACTACTCAACTTATTAAAAACAACGACTTTCATAAAATTTTA
GGCAAATGNATGGTCTGNAGGATCTTGAGTGAGGTAACCCAATCACAAAAGAACAC
TCATGGTATGCACTCACTGATAAGTGGCTATTTGTCTATGGAGTGATTTAAAAGGGA
AGAAGACACATAGCTTTTTGTGTGTATAATATTAAGATGGAAATTTGCCAGTGCTGT
TTGGCTTATGAGTGAATCTTGTTTCAGTGGATTACCGGAAGAAAATAATAAGTGAAC
TGTAGGAAGAAGTAGTTAATCAAGGTGACAAAGTATCCTGACACATTGGGAAAAGA
CCACAGTCCAGGAAACTGAGTCTTAAGGATTCATATTAACTCCAGTTCCCCATGTGC
AGCTCTGAGACTTTGGCAGATCAGACACTTAACTTCACCAGCTTCCTACACAGAGCA
GTTACTATCCTTGCCTTCACACATGGAGTGTGCCATTAAGTGCCTGAACATGAGTCT
GACTTGTTAATAATCTTTAAAATCCAATTGTGTGTAAAGTATGTGACCAAAGAGCAT
GGTCATGCTATTAACCTTTGATGTTCTATGGACTCTTAATTTTATGGTAGAAATGTCA
ACAAGCTTGTGGAGGCTGGAAGATACAAGGCTTAAGAGGATGGCCTTTCAGTTTTG
AAAGTAATTCAGTATGTGTTCTGGCATCCCTTTTCCTAAAGCAATTTAACCCCCCAA
GTAGGCATAATTTTAATGCTTACTTCATCAGAATATATCTAATTGACTCTTCTAAAAA
GACTTTGGTATGCATAGGATCTAAATGTAAATGTGATTTACTGACATAATAAATAGG
AGAAACTGAGCTAGAATAGGTATAAAATATGTGCTGGCTTTCTAATAGGTCTTATAG
GTTATATAAGAGGTGGGAAAGGAATATTTGAAACATCTAGAAGTAAAATGATCCTG
AGTAGCGATCCTGGGAAAATACGTACTCTAACACACTGCAATCATCTCTGTGGTT
TGCTGGAGCTGAGGTCTGGAAGGCTCGACCTTGGTTAGAAATAACCTACCGAATAC
AGAGCTATGACGTTAGTCTGGAAGGAGCTTTGGAAGAATGACAAGCTGTAGCTGCC
CAGAACATACTAGATGCCATATTTCCAAGGCAAGTGTCCACATGCGGACATCTTAAG
AATATGGTTGTCTCTGCAGTGCTAAGGACCTTGTTCGTGCCACACAGGTCTCCAGGG
TTAGTGCTAACTCTGACTGCTTGACTCTTTAATTCTCCCTTGATCATTAATGACTAGA
AATCACTTGGTGATTAGCAACTGGATATGGAATATTACTTAATTTGTACCCAAGCCA
GGCCACCTCAGCTTTGGCAGCTCCATTCATTCTGTGGAGCCCAGTCACGTGGGTTTG
AATCAACTGTACTGTTTCTACTTACAAGACGCATTACCTGAGATGAGTCATTTTTCTT
CACAAGTCTTTTTAGAAGAGTCAATTAGACATATTCTGATGAAGTAAGCATATAAAG

TGAGAGCAGCATGAATGTGTTCCATGTATGCTCATGGATGCTATTATAATGTGGAAA
TAAACTGACTTTAAAAAAAAAAGCTTATGATACTTGTCACAGAGTAAATCTTCCATA
AATATCATCTGCATTTATAAATTATTTTCATAATCCATCAATTAAAAACCTTTAGAAA
TTTTGTTAACACAAAGATCCCTAGGCCCCTGCCCTAGGATGGTCTGTATGGTGGGCC
TGAGAGATGGAGCTTAAGAACTTACTTGCTCCAGGAGCACATCTTCAGAACATCTGC
CTCAAAACATTTATCCCAAATGCTCATCAAAGGCTCACTCACATGTGCTTCAACCAC
AGGGATTAAACAGTCATTTTAGTCACATTTCTCAAACGGTGGAAGCCTGCTAGAGGA
ACAGGATGTATCAGGATAACATCCAACCTTACAAAAGGATGTCATAACCCTCACCA
CAACAAACAACAACGACAACAAACCCATAAAAATTATCACGGCAAATGAACTAAGC
CATATGCAGAAAAAGTATTATATGTTCTCATTGTGGGGTGTTTTTCCTTAATAGTCAA
ATATGCAGAATATAGACAAAGATGGTTTATGCAAGTGGGGATGGCGAAGGATACTT
GTAGATTAGAGGACACAAAGCAACAACTACAGAGTGAAGTAATCCAGAGACTTAAT
GTATAATATGAGGACTGTATTTAATAATTCTATTTAAGATACACAGCAAACGAGTGT
ATCTTACTAACACACACTTACATAGAGAGAATAAAGTGATAGATACGTTTGTTTT
ATCTTCATGTAGCTGATAATTTCATATTGTACACCTCAAACATAGATAACCAACAAA
GAGGAAGAGGATAGGTGCCTCTCCCAGGGCGGAAGAGTACATTCGAAAGTCAGACA
CCATTGTGTAGATGTACCACATGGAGGAGCTAGAGAAAGTAGCCAAGGAGCTAAAG
GGATCTGCAACCCTATAGGTGGAACAACATTATGAGCTAACCAGTACCCCGGAGCT
CTTGACTCTAGCTGCATATATATCAAAAGATGGCCTAATCGGCCATCACTGGAAAGA
GAGGCCATTGGACTTGCAAACTTTATATGCCCCAGTACAGGGGAATACCAGGGCCA
AAAAGGGGGAGTGGGTGGGCAGGGGAGTGGGGGTGGGTGGATATGGGGGACTTTT
GGTATAGCATTGGAAATGTAAATGAGTTAAATACCTAATAAAAAATGGAAAAAAAA
AAAAAAAAAAAAAAAAAGGAAGGTCAGACACCTCACTTCACTGCTATCTCAACTTG
CAAACAGAAGGGGAGTCACAAACCCAGGACAAACCACAGTGATTGAAGCGTCTTTG
AATGTTATTGCTGTTGTTGTTACCACCATCATTAGCATATATTCATTGTGAAAACTTA
CGGGGTCTATGACATGTTTTTTTATTCAAGTATATCACATGCTGTCAGCATATTTGGC
ACCACTACCAGCCCCAGCCCCCTTTGCCCCGCCCCCAACACACACACACACACACAC
ACACACACACACACACACACACACACACACACACACACCTTTACCTTCTCCTGGGCA
TCATCTGCTCACTCACCCACCCAAGCTTAATCCTTTTCCTTCCCTGCAATAGTACCTC
TCCTATTTTTATGTCTAGGTTCCCCCTCCCCCTGTTAGGAGATGGGAGAGGTCACGA
AAGAAAGGAATTTGTAGCCCTTGAGCCAGCCCGGGCCACAGAGCCTGCCACCAGAC
AGGAAAAGCCCAGGGCTTACCAGCACAGGAGGAGCAAACTCGCAGGCGAGCCTGG
GTTGGCGCTGGTGGTCCCGGGTCGATGGCCCGCCCATTCCCAGAAGCCGAGGCTATA
GCTGCGTCACCTGCCCCGCCCTCCTCCCGAGTGAAGACCCCTAGAGGCTGAGCAGAC
CCCAAAGGCGGTGCAATTCCATTGGCCCAAGGCAGAGGTGAGCGGCTGCTAATCCC
CTCGGGAAGTGAAGGGACCCAGAGAGTCTGGTAGATGTGGGAGCTGGGGTTCAGGG
CGAGACAGAGGGTGGGATGGGCAGAAGGGTCCAGGAAAGGAAAGTACTGGAGGGG
AGTTGGGACAAAAGCAGCGACCAAGGGAACATCGCTTCAGTGACTGAAGCCAGGCA
AAAGGAGCGGGAAGGATTATATGTAGCCTGGGACGCTTTCATAAACACTGATGACG
TGTTTGTGCAAAGCAAGCAATTTGAGGAGAAACGCCTGGGACGTCGGAAAGAAGGA

GTGATCGATTAGTACTTGTAAGTTTAGGTGAGTTTGAGAACTAACTAACCTATACTA
TTGAGGGAGAAGGAAGAGCATTCCAGCAGCAGCAGCAGCAGCAGCAATCAGATAA
AGGAAAGCTTTGGTTAGTTTGGAAATGTATGATACCATTAAAATAACAGAAGCGCCT
CCAGTTCTCTGAAGAGTCAGTCCCCCAGCTAGTGAAGACTAAGCCTACTAAGCCTTT
TGCTCCCGTTGGAAGCAAAGAACGTTCCTTCAATCAGGTGAAGGCTCTCCTCAGAAG
ATTTCCTGTCTCTGCTTATGTTACAAGAGGATTCAAAAGCAAGACAGAAGAGCTCAG
GTATTGCCAACTCTTTGTTAAATACAGTTTGAGGCTTAAGTGTACGGGAACTCATG
TGGTATTCATTTACGGCTCTCTTCTCTTATAACTAACTCTTAAGGTGCATATAGTCTC
TTCTGTTTCCCAGCTACCTTGTACCATCTTTGTTTATCTAATAATAGCAAGCTCATCT
GCTTTTTAATCATCACGCAGAGAGTATTCAAAAATATTCAGTGATGTAACAGTGACA
GTGTAGGCATAGAAGTAATCATTAGTAAATCTTAATTTGGGTTAAACTCATTCATAA
CAGCTCCAGGTTGGGAGGGATCACTGAGCCTTCGCCACGTGCGGGTTAAAGATATTT
TCTAACAAGAGAAGCAGAATTCTTCCTTGGCCATGCTCCCCATCACTGTGTCAGTAA
GCAGAGGGGTGTTTCCAAGCAGAGAAAGAGCAGACAGTGTTATGCCTGCAAAGTCA
GAGACTCAGCCCTCCCAGCTGGTCAGTTACTGTCCTCCCGGTCATTAGTTGGCTCTG
AAAAGGCCCATGTGTCCTTATTGGCAAGGACTTGCAGACATGCTAGAAAGAAATTT
GACCTTTTTTCTAGTGGGTTATTACAGCTGTAAAGTATTTTGGAAGGTTAAGCCA
AATAAATAAAACACATATTAAATAATACAATGTTACAAAAATTGATCATATAAAGA
AGTACATTCATAAATGCAATGTGAAAAATATATATAATTTTTATCTATTTACTGGTGC
AAAGTTTTCTAAATTGCACATGTACTATTTTTATATTTATAAAAATATTTTTAAAATG
TATATAAAAGTGTAAAAGGCTCTTGGTCAAACAAGAGAGTTAAATTTACAAACTTTA
ATTGTCCCGATAACATTATTATGATCTCTAATGACAGGGATCCTGCTTTTCATTGGGA
AATGAGAAGCTATGAAGATATGTTTACAATAATAAGCCCATTTAGTGATAAAGTCCA
ATGGGAAGCTAGCACACACTGGTTTATAAAGAGAACAGTTTCCTGAGTCTATGCAA
GTTTACACTCTAGGGAATAAGAGTTCCTCTTTCTCCAGATTTCACTAGCATTTGTTGT
CATCATTTATCTTCTTGATGATGAGCATTATAAGTGGAATAAGATAGGATCTCAAAG
GAATGTCAATTTGGATGCCCTGAACAATCTTTCAGGTCTTTCTTTCAGTTCACTAGTC
TATTCATTTATTGGATAATTGGGGGGATGGTGGTAATTTTTTTGCAGTTCTTATGGAA
TTCCAAAAAACAAAAAACAAACCAACCAACCAAAAACCTCTGAAACTAGAACTACC
AATCCATTACTGGGTATGTAACAAAGAGAAATCTGCACAGAATTTATTGCTACATTG
TTCATTATTCACGACAGCCAAGAATGTGGAACCAACTTACGTAGCCGTCAAAATATG
AACGGATAAAGAAAATGTGGAAATGTGTACAACAGAGTCCCATGTGGCCATAAAG
AGTGAAATCATGACATATGCAGGAAATGGATGCAACTGGAAATCAATTGGGCTAAT
CAAAACAAGACAGACTCAAAAAGGAAACACCGTGTAGCTTCTCTGACAAACAGAAG
CTAGATTTACACTTGTACGTGCGCATGTGTGTTTAGAATTTTATTTAGTTATACACTA
TTCTAATCTGTGAGTGTGTATAAAGGCATGCATGTAAAGCAAAAACAAGCTAGCTG
GGGTGGGTAGGAGAGAAAGCAATGAGAGGAGTTAATAAGAACGAAGCATAGTAAC
ATAGGTGCCAGGATGAAATGCATTAATTTGTATGCTAACTAAACCACAGACAGGAG
GCACACGTTCAAACCAGGGTGAAATCCCAGCACAGAGAAGGGGAAGTAGACACAA
AGTTTCGCCACTAACCAAGAAGCCATTTGCAGTTGCTGCCTGCTGGGAAGGGGCGTT

CCAGTTTTCTCCAGTCTGACACTGTGTATAACAACCAGTTGACAATACAAAGTTGGC
ATGATGGATGGTTTTGTGCTATTTTTCATTTTTTTCTTACTGTTTTGTTGTTGTGGT
GGTTGTTGTGGTGGTGGCTGTGGTTTTCATTTGTTTCTTTTGAGAGAGAGAAGGAAC
ATGAAATTGGGTGGGTAGGAAGCTGGAAACGATCTGGAAGAAGTTGGGGAAAGAG
AAAAATTGTATGGAGCATATTTAAACAAACAAACAAACAAACAAAAGGTTCATTTT
GCCACAAAAAGGTGTGAATTAAATTAACCAGTTACGACTCTTAAAGAAAATATTCCC
AATTATTCCCAGAGTTGCTATGTATGCTGTGCCTAGGACTTTGCTTGAACTGGCCCTA
TAACTCTGGTGTGGTGTCTTTTCAGGATGCAGAAGAGAGGCAGGGAAGTCAGCTGCT
TGCTGATCTCCCTCACTGCCATCTGCCTGGTGGTCACCCCTGGGAGCAGGGTCTGTC
CTCGCCGATGTGCCTGCTATGTGCCCACAGAGGTGCACTGTACATTTCGGGACCTGA
CCTCCATCCCAGACGGGCATCCCAGCCAATGTGGAACGAGTCAATTTAGGGTGTGTG
GACCTTGCCTGATCTCCTTCTCAGAGAGGGACCACTGATTTTCCTGGTACTTTGCCCC
CCAAACACCTGTGATTACTTTTAATAGTTTTCTTCTAAAATGGGTTCATACAAACCTT
ATATTGTGGAGACAATGAACATTTTATCCCAATAGTCTTTTACTAGAACTTGAAGCC
CCTCTTAGTTGTTTGGGAGCCTCATAATTATGGGGCAGCTTTATTCTGAATGAATTTT
AAATGAAAAAGATACAGTTTCTGTTAACAATCATTATGATACCAAGGAAGAGGAAT
TGTCATTGAATATTTTAAAAAAGCATTTCTTTTGCAATTTATAAATACCCATTACAAA
ATGGCTTACTTAAAATACTTGCCTTACTAAATCTGACAAATTATGGTGATATTTTGAA
GGTTTATGAAAATTTGTTTATGTGTATAAATGCACAAGAAATGGGATATGCCATCAC
CTATGTGCCATTAGTGAGCATGTACAGTATGCCAAACACTATTGTTCACGTTTGGAG
GAAGTAATGGGGGTGGGGGAGCAACAAGGGTTATAACCGTATACCCAGTGCCTTGG
AAGCGATTGCAAACAGTAAAGACTGACATTGTGTTCTCCCTATGAGGGAGGGGCCTT
GGGCTGAGCACTTTGCAATGAGCATTTGCTCATTGTGCTGGCAGGTTTTATGATAAC
TTGACCCAAGCTAGAGTCACTGGAGAGGAAGGAACTTCAACTGAGAACATGCCTGA
AGAAGATCAGATTATAGGCAGGCCTGTGGGGCATTTTCTTAATTAGTGATTCATGGG
GCAGGGCCCAGTCCATTGTTCGTGGTACCATTTCTCAGGCACTATTAAAAAAAAAAA
AACAGGCTGAGCAAGTGTCAAGGAGCAAGTCAGTGAGCAGCAGCCCTAATGATCTC
TGCATCAGCTCCTGCCTCCAGGTTCCTACCCTATTTGAGTTCCTGTCCTAGCTCCCTA
CAGTGATGAACAATGATGTGGAAGTATAAGCCAAATAAATCCTTTCTTCCCCAACTT
GCTGTTGGTCATGATGTTTCATCACAGTGATAATAGTCCTCATGAAGATGCTGGTGT
TTATAACACCTTTGGACTAAATTCTGTTATCTATAGCTGAGGAAAATGGAGCATAGA
AAGTCTCCAGACTACACCAGAGTGTAATCTGGGCCTGAGCTTAGAATCACACCCAC
GTGCACTCCACTGCCGGGGCTTCTTAACCGGAACACAGTTGTAAAAGGGAATTTTCT
GTTTGTTTCCATTTTGACATGTGGACTTTAATTGACGATTCATCTGAAGCTGAAAATG
ATTTTTTTCCAGGTATAACAGCCTCACTAGATTGACAGAAATGACTTTTCTGGCCT
GAGCAGACTGGAGTTACTCATGCTGCACAGCAATGGCATTCACAGAGTCAGTGACA
AGACCTTCTCGGGCTTGCAGTCCTTGCAGGTGAGATAGGTAGAGGGTGATGGAGGC
TGAGAAGAGAGGTGCAACTGTGGGTTATACCCAAAAGCTGCTGATTCCCGTGGGAG
ACATTCTATAAGCATTCTATAAACTAGAGGCAGATATCAAGGAAGGATTTCAATTGT
AATGCAATTTTATGAGAAAATTTGAATATTAAGAAAATGCTGGGGAAAATGCTTAC

ACAATTGCGAGGACCTAATTTAGGATCTCCAATAGCCACATAAAAAGCACAGCATG
GCGGCAGACACCTGCAATTCCTGTCCCTGGAAGCACCTGTTCAGAATCCCAGAGACT
CATTGGCCAAACACTCTATTCAATCAATGAAGTCCATATTCAGTGACAAAACTTGAC
TCAGAAACTAATGTGGAAAGCATCAGGAAGACAGCCAACATCTGGTCTCTACTCAT
GCATGAATAAGGGATCCCAGAGAGAAGGGAAGAAAAAGGAAGGAAGGAAGGAAG
GAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAGAGGGAGGAAAGGAGGG
AGGGAAGGAAGGAAAGGGAAAGGAAAAAGAGATGGGGAGGGAGGGAAGGAAAG
GAAAGGGGGAGAAGAAGAGAAGAAAGGAAATAAATAAATTTTCAGGGATTATT
ACACCTTTAAATTTTATCCATAAAAGGTCATTTCCACCTGTTTGTCTGGAAGTAGAGT
GGGATCCCTTATATAAGGGCAGTCTTTAACATAGTAGCATTTTATAAACCATTACAA
ATTTTGAGTTTTCTCTACTTTTTATCCTCTACCATCTTCAAACTGAAACTACAATTATT
CCCACAAATGAAGAAAATGCTGTAAGAGTTTTCACACACCGAAGTGGGAAACTTAA
GGATTAGACAAGTCTAACAATGAGAATGGGGAGAACAAAAAGAGACTGCACAGGG
AGCCCTTTCTCTGCTTATAATCTTGACACTTGAGAAGCTAATTGACGCTGCATGACTA
CTCAACTCTTTAAGCAAACAATGCTGTTGTTCATGAAAAGCACAATAAAGTACATAT
GTCCCATAATATTCATCAAAATTTGCATGCAGCACATAATAGCAATCAAAGCAATAA
CACCCACTGTTCACAGAGACTTTAAACATGAAACTGGAACTATGTCTAGTGTTTTGA
CTTAGGGTACATAGTATGCTGTGTCTGTATGTACCAATGTTGATTAGGTCATCAGA
CAGCATTTGGAACATGTATCTTCAGGAGGAATCATTCATGTATCCTGCATGAAATTC
TCCACCTATGTTTATTCTCTTAGCCAGGTTTTTCTCTGATGGAGAAACATTGGGTTTG
AGGTTTTACTCCCAGGTAACATTTAGGGAAAGCTGTCTATGTTCTCAGTTTGGCTTT
TATTTATGAGGGATGTTGGTATTCCAGAAAATTCTCTTTTGAAGAGATTACAATTTA
GGTCAAAACAGAAAAATATGTAAAAGTTATTGTTTTTATTAGTATTTCATGTTCTTT
TCTTTTTTAAAAATGGTATGCTTAGAACTAATTAAGATTAGATTAGATTAGATTAGA
AAATAATCAGAGAGGGATTTGATGAATGCTAAAGCATCATGAAAAATTCAAAATTT
TTTGCTTCTAATTCAGAATCAATTAAATTCATATTACTATAAAAGACAGCACGCCAG
ATGTGTGCCAGCTGAGGAGTGGATAAACTGTGTAACGTGAGTGCTATGTAGAAACA
GAAAGGAGTGAAGGGTTGATGTGCGCTGCAACATCTTGAAAACATTCGGCTACATG
ATGGAAGCCAGGCACAAAAAGCCACATATTGCATGGTTATGTTTATATGAAATGTTT
AAAATACATGGATTCTTAGCAAACAGAGTAAGATGTTACTTAGGGTCAGGAAAAGA
TTAAAAAAAAAAAAACTATTGATGTGGAATGATCTTAATTTGGGGAAAAGACAATT
TCCTAAGACGAAATAGTTGAGGTAGATATAGTTATATCCCTGTGGATATTGTAATAA
ACCAGCATGCTGTGCTCTGAGAAGGGCCTAATGAAGGGGCAGGAGGAAGTGAAATG
AGATGGTAGAAAGGAAAGTCATATACCATGGCTTCTCTCGTGGGTGGAATCTAGAT
ATGTTAATATATTGACATAAAGGAAGGAATTGTTAGGGAAGGATCAAAACCAACA
GGAGTGAGGGAGACAATAGGAACCAATGAGAGGCAAAGTTCATGGTCAATGTGTGT
GGAGACACCATAATAAAACTCCTTTTTTGTTTGCTAACTAAAACCACTAAAATCTAA
AAACAAAACATTTTTGCACAAGAATTATTTATTATTCAATAAAGATGTTTAAATGGG
GGAAGTTGAAGTTCATTGATAGTCTCATAAATCTTAAATGTATTTAAACTGCTTTTTA
CGTTTTTTATTATTAATTACTCTTGCTGTCATTATTATCATCATCATTATCGTCATCAT

CATCACTAATGCTTTTCACCATACACAAATGTAGGCAGAAGAGTGTAATCCACTTAG
TGAGGCAATCTTGGAGAGGGAAAGGAAGCGGATGCGGGGCAGAGGCACACAGGAG
GACAGTGAGAGGGAAATGAACAAGAAAAAATGTGGACACATGCACAAAAATTCCA
TAGTCCACTACATTACTTTGTATTCTAATATTAAGAAAATAATAAACCCATTTCTGTG
CACTTATCACCCAGGCTCAACAGTTATCTTGGCCACAGATCCTGTCTCACTGCATCCT
GTCCACCTGAGTCCACTTAGCGTTCTGAATCCAATCCAGGGCATGATGCTTACTCCT
ACACAGAACTAAAGATTAAAGAGAGTTTAAAAGTAACCATGACATCTCTCTGTTCCT
TTAGCGATAAGTTCTTAATATTTATGGCTGCTTGTGTATGTTCTAATTTCTCTAATATT
GTCACATTAGTTGGCAACTACTTTGTTTGAATTGAGTTGGAGTTAAGGTCCCATAG
GATTAATCTCAACATATTTCTATATTTATAAACTTTTCTCTCTTGTGAAAGTTCCTTT
GAGAAACAAATATGCCCATATCTTTCTTTACAGGTCTTAAAAATGAGCTATAACAA
AGTCCAAATAATTGAGAAGGATACTTTGTATGGACTCAGGAGCTTGACCCGGTTGCA
CCTGGATCACAACAACATTGAGTTTATCAACCCCGAGGCGTTTACGGACTCACCTT
GCTCCGCTTGGTACATCTAGAAGGAAACCGGCTGACAAAGCTCCATCCAGACACAT
TTGTCTCTTTGAGCTATCTCCAGATATTTAAAACCTCCTTCATTAAGNACCTGTACTT
GTATGATAACTTCATTGACCTCCCTCCCAAAAGAAATGGTCTCCTCTATGCCAAACC
TAGAAAGCCTTTACTTGCATGGAAACCCATGGACCTGTGACTGCCATTTAAAGTGGT
TGTCCGAGTGGATGCAGGGAAACCCAGGTAACTATCTTGTTTGTTTGTTTCTTTTTTT
ATARKACGTATTTTCCTCAATTTCATTTAGAATGATATCCCAAAAGTCCCCCATAACC
TCCCCCCCACTTCCCTACCTACCCATTCCCATTTTTTGGCCCTGGCATTCCCCTGTACT
GGGGCATATAAAGTTTGCGTGTCCAATGGACCTCTCTTCCAGTGATGGCCAACTAG
GCCATCTTTTGATACATATGCAGCTAGAGTCAAGAGCTCTGGGGTACTGGTTAGTTC
ATAATGTTGTTGCACCTACAGGGTTGAA (SEQ ID NO:20)

Figure 51

MPKRAHWGALSVVLILLWGHPRVALACPHPCACYVPSEVHCTFRSLASVPAGIARHVE
RINLGFNSIQALSETSFAGLTKLELLMIHGNEIPSIPDGALRDLSSLQVFKFSYNKLRVITG
QTLQGLSNLMRLHIDHNKIEFIHPQAFNGLTSLRLLHLEGNLLHQLHPSTFSTFTFLDYFR
LSTIRHLYLAENMVRTLPASMLRNMPLLENLYLQGNPWTCDCEMRWFLEWDAKSRGIL
KCKKDKAYEGGQLCAMCFSPKKLYKHEIHKLKDMTCLKPSIESPLRQNRSRSIEEEQEQ
EEDGGSQLILEKFQLPQWSISLNMTDEHGNMVNLVCDIKKPMDVYKIHLNQTDPPDIDIN
ATVALDFECPMTRENYEKLWKLIAYYSEVPVKLHRELMLSKDPRVSYQYRQDADEEAL
YYTGVRAQILAEPEWVMQPSIDIQLNRRQSTAKKVLLSYYTQYSQTISTKDTRQARGRS
WVMIEPSGAVQRDQTVLEGGPCQLSCNVKASESPSIFWVLPDGSILKAPMDDPDSKFSIL
SSGWLRIKSMEPSDSGLYQCIAQVRDEMDRMVYRVLVQSPSTQPAEKDTVTIGKNPGES
VTLPCNALAIPEAHLSWILPNRRIINDLANTSHVYMLPNGTLSIPKVQVSDSGYYRCVAV
NQQGADHFTVGITVTKKGSGLPSKRGRRPGAKALSRVREDIVEDEGGSGMGDEENTSR
RLLHPKDQEVFLKTKDDAINGDKKAKKGRRKLKLWKHSEKEPETNVAEGRRVFESRRR
INMANKQINPERWADILAKVRGKNLPKGTEVPPLIKTTSPPSLSLEVTPPFPAVSPPSASP
VQTVTSAEESSADVPLLGEEEHVLGTISSASMGLEHNHNGVILVEPEVTSTPLEEVVDDL
SEKTEEITSTEGDLKGTAAPTLISEPYEPSPTLHTLDTVYEKPTHEETATEGWSAADVGSS
PEPTSSEYEPPLDAVSLAESEPMQYFDPDLETKSQPDEDKMKEDTFAHLTPTPTIWVNDS
STSQLFEDSTIGEPGVPGQSHLQGLTDNIHLVKSSLSTQDTLLIKKGMKEMSQTLQGGNM
LEGDPTHSRSSESEGQESKSITLPDSTLGIMSSMSPVKKPAETTVGTLLDKDTTTVTTTPR
QKVAPSSTMSTHPSRRRPNGRRRLRPNKFRHRHKQTPPTTFAPSETFSTQPTQAPDIKISS
QVESSLVPTAWVDNTVNTPKQLEMEKNAEPTSKGTPRRKHGKRPNKHRYTPSTVSSRA
SGSKPSPSPENKHRNIVTPSSETILLPRTVSLKTEGPYDSLDYMTTTRKIYSSYPKVQETLP
VTYKPTSDGKEIKDDVATNVDKHKSDILVTGESITNAIPTSRSLVSTMGEFKEESSPVGFP
GTPTWNPSRTAQPGRLQTDIPVTTSGENLTDPPLLKELEDVDFTSEFLSSLTVSTPFHQEE
AGSSTTLSSIKVEVASSQAETTTLDQDHLETTVAILLSETRPQNHTPTAARMKEPASSSPS
TILMSLGQTTTTKPALPSPRISQASRDSKENVFLNYVGNPETEATPVNNEGTQHMSGPNE
LSTPSSDRDAFNLSTKLELEKQVFGSRSLPRGPDSQRQDGRVHASHQLTRVPAKPILPTA
TVRLPEMSTQSASRYFVTSQSPRHWTNKPEITTYPSGALPENKQFTTPRLSSTTIPLPLHM
SKPSIPSKFTDRRTDQFNGYSKVFGNNNIPEARNPVGKPPSPRIPHYSNGRLPFFTNKTLSF
PQLGVTRRPQIPTSPAPVMRERKVIPGSYNRIHSHSTFHLDFGPPAPPLLHTPQTTGSPSTN
LQNIPMVSSTQSSISFITSSVQSSGSFHQSSSKFFAGGPPASKFWSLGEKPQILTKSPQTVSV
TAETDTVFPCEATGKPKPFVTWTKVSTGALMTPNTRIQRFEVLKNGTLVIRKVQVQDRG
QYMCTASNLHGLDRMVVLLSVTVQQPQILASHYQDVTVYLGDTIAMECLAKGTPAPQI
SWIFPDRRVWQTVSPVESRITLHENRTLSIKEASFSDRGVYKCVASNAAGADSLAIRLHV
AALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRWVLGDGTQIRPSQFLHGNLFVFP
NGTLYIRNLAPKDSGRYECVAANLVGSARRTVQLNVQRAAANARITGTSPRRTDVRYG
GTLKLDCSASGDPWPRILWRLPSKRMIDALFSFDSRIKVFANGTLVVKSVTDK

DAGDYLCVARNKVGDDYVVLKVDVVMKPAKIEHKEENDHKVFYGGDLKVDCVATGL
PNPEISWSLPDGSLVNSFMQSDDSGGRTKRYVVFNNGTLYFNEVGMREEGDYTCFAEN
QVGKDEMRVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTWLSPTN
KVIPTSSEKYQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQPPKING
NPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVLPAPYYGNRITVHGNGSLDI
RSLRKSDSVQLVCMARNEGGEARLIVQLTVLEPMEKPIFHDPISEKITAMAGHTISLNCS
AAGTPTPSLVWVLPNGTDLQSGQQLQRFYHKADGMLHISGLSSVDAGAYRCVARNAA
GHTERLVSLKVGLKPEANKQYHNLVSIINGETLKLPCTPPGAGQGRFSWTLPNGMHLEG
PQTLGRVSLLDNGTLTVREASVFDRGTYVCRMETEYGPSVTSIPVIVIAYPPRITSEPTPVI
YTRPGNTVKLNCMAMGIPKADITWELPDKSHLKAGVQARLYGNRFLHPQGSLTIQHAT
QRDAGFYKCMAKNILGSDSKTTYIHVF (SEQ ID NO: 21)

Figure 52

ATGCCCAAGC GCGCGCACTG GGGGGCCCTC TCCGTGGTGC TGATCCTGCT
TTGGGGCCAT CCGCGAGTGG CGCTGGCCTG CCCGCATCCT TGTGCCTGCT
ACGTCCCCAG CGAGGTCCAC TGCACGTTCC GATCCCTGGC TTCCGTGCCC
GCTGGCATTG CTAGACACGT GGAAAGAATC AATTTGGGGT TTAATAGCAT
ACAGGCCCTG TCAGAAACCT CATTTGCAGG ACTGACCAAG TTGGAGCTAC
TTATGATTCA CGGCAATGAG ATCCCAAGCA TCCCCGATGG AGCTTTAAGA
GACCTCAGCT CTCTTCAGGT TTTCAAGTTC AGCTACAACA AGCTGAGAGT
GATCACAGGA CAGACCCTCC AGGGTCTCTC TAACTTAATG AGGCTGCACA
TTGACCACAA CAAGATCGAG TTTATCCACC CTCAAGCTTT CAACGGCTTA
ACGTCTCTGA GGCTACTCCA TTTGGAAGGA AATCTCCTCC ACCAGCTGCA
CCCCAGCACC TTCTCCACGT TCACATTTTT GGATTATTTC AGACTCTCCA
CCATAAGGCA CCTCTACTTA GCAGAGAACA TGGTTAGAAC TCTTCCTGCC
AGCATGCTTC GGAACATGCC GCTTCTGGAG AATCTTTACT TGCAGGGAAA
TCCGTGGACC TGCGATTGTG AGATGAGATG GTTTTTGGAA TGGGATGCAA
AATCCAGAGG AATTCTGAAG TGTAAAAAGG ACAAAGCTTA TGAAGGCGGT
CAGTTGTGTG CAATGTGCTT CAGTCCAAAG AAGTTGTACA AACATGAGAT
ACACAAGCTG AAGGACATGA CTTGTCTGAA GCCTTCAATA GAGTCCCCTC
TGAGACAGAA CAGGAGCAGG AGTATTGAGG AGGAGCAAGA ACAGGAAGAG
GATGGTGGCA GCCAGCTCAT CCTGGAGAAA TTCCAACTGC CCCAGTGGAG
CATCTCTTTG AATATGACCG ACGAGCACGG GAACATGGTG AACTTGGTCT
GTGACATCAA GAAACCAATG GATGTGTACA AGATTCACTT GAACCAAACG
GATCCTCCAG ATATTGACAT AAATGCAACA GTTGCCTTGG ACTTTGAGTG
TCCAATGACC CGAGAAAACT ATGAAAAGCT ATGGAAATTG ATAGCATACT
ACAGTGAAGT TCCCGTGAAG CTACACAGAG AGCTCATGCT CAGCAAAGAC
CCCAGAGTCA GCTACCAGTA CAGGCAGGAT GCTGATGAGG AAGCTCTTTA
CTACACAGGT GTGAGAGCCC AGATTCTTGC AGAACCAGAA TGGGTCATGC
AGCCATCCAT AGATATCCAG CTGAACCGAC GTCAGAGTAC GGCCAAGAAG
GTGCTACTTT CCTACTACAC CCAGTATTCT CAAACAATAT CCACCAAAGA
TACAAGGCAG GCTCGGGGCA GAAGCTGGGT AATGATTGAG CCTAGTGGAG
CTGTGCAAAG AGATCAGACT GTCCTGGAAG GGGGTCCATG CCAGTTGAGC
TGCAACGTGA AAGCTTCTGA GAGTCCATCT ATCTTCTGGG TGCTTCCAGA
TGGCTCCATC CTGAAAGCGC CCATGGATGA CCCAGACAGC AAGTTCTCCA
TTCTCAGCAG TGGCTGGCTG AGGATCAAGT CCATGGAGCC ATCTGACTCA
GGCTTGTACC AGTGCATTGC TCAAGTGAGG GATGAAATGG ACCGCATGGT
ATATAGGGTA CTTGTGCAGT CTCCCTCCAC TCAGCCAGCC GAGAAAGACA
CAGTGACAAT TGGCAAGAAC CCAGGGGAGT CGGTGACATT GCCTTGCAAT
GCTTTAGCAA TACCCGAAGC CCACCTTAGC TGGATTCTTC CAAACAGAAG
GATAATTAAT GATTTGGCTA ACACATCACA TGTATACATG TTGCCAAATG
GAACTCTTTC CATCCCAAAG GTCCAAGTCA GTGATAGTGG

TTACTACAGA TGTGTGGCTG TCAACCAGCA AGGGGCAGAC CATTTTACGG
TGGGAATCAC AGTGACCAAG AAAGGGTCTG GCTTGCCATC CAAAAGAGGC
AGACGCCCAG GTGCAAAGGC TCTTTCCAGA GTCAGAGAAG ACATCGTGGA
GGATGAAGGG GGCTCGGGCA TGGGAGATGA AGAGAACACT TCAAGGAGAC
TTCTGCATCC AAAGGACCAA GAGGTGTTCC TCAAAACAAA GGATGATGCC
ATCAATGGAG ACAAGAAAGC CAAGAAAGGG AGAAGAAAGC TGAAACTCTG
GAAGCATTCG GAAAAAGAAC CAGAGACCAA TGTTGCAGAA GGTCGCAGAG
TGTTTGAATC TAGACGAAGG ATAAACATGG CAAACAAACA GATTAATCCG
GAGCGCTGGG CTGATATTTT AGCCAAAGTC CGTGGGAAAA ATCTCCCTAA
GGGCACAGAA GTACCCCCGT TGATTAAAAC CACAAGTCCT CCATCCTTGA
GCCTAGAAGT CACACCACCT TTTCCTGCTG TTTCTCCCCC CTCAGCATCT
CCTGTGCAGA CAGTAACCAG TGCTGAAGAA TCCTCAGCAG ATGTACCTCT
ACTTGGTGAA GAAGAGCACG TTTTGGGTAC CATTTCCTCA GCCAGCATGG
GGCTAGAACA CAACCACAAT GGAGTTATTC TTGTTGAACC TGAAGTAACA
AGCACACCTC TGGAGGAAGT TGTTGATGAC CTTTCTGAGA AGACTGAGGA
GATAACTTCC ACTGAAGGAG ACCTGAAGGG GACAGCAGCC CCTACACTTA
TATCTGAGCC TTATGAACCA TCTCCTACTC TGCACACATT AGACACAGTC
TATGAAAAGC CCACCCATGA AGAGACGGCA ACAGAGGGTT GGTCTGCAGC
AGATGTTGGA TCGTCACCAG AGCCCACATC CAGTGAGTAT GAGCCTCCAT
TGGATGCTGT CTCCTTGGCT GAGTCTGAGC CCATGCAATA CTTTGACCCA
GATTTGGAGA CTAAGTCACA ACCAGATGAG GATAAGATGA AGAAGACAC
CTTTGCACAC CTTACTCCAA CCCCCACCAT CTGGGTTAAT GACTCCAGTA
CATCACAGTT ATTTGAGGAT TCTACTATAG GGGAACCAGG TGTCCCAGGC
CAATCACATC TACAAGGACT GACAGACAAC ATCCACCTTG TGAAAAGTAG
TCTAAGCACT CAAGACACCT TACTGATTAA AAAGGGTATG AAAGAGATGT
CTCAGACACT ACAGGGAGGA AATATGCTAG AGGGAGACCC CACACACTCC
AGAAGTTCTG AGAGTGAGGG CCAAGAGAGC AAATCCATCA CTTTGCCTGA
CTCCACACTG GGTATAATGA GCAGTATGTC TCCAGTTAAG AAGCCTGCGG
AAACCACAGT TGGTACCCTC CTAGACAAAG ACACCACAAC AGTAACAACA
ACACCAAGGC AAAAAGTTGC TCCGTCATCC ACCATGAGCA CTCACCCTTC
TCGAAGGAGA CCCAACGGGA GAAGGAGATT ACGCCCCAAC AAATTCCGCC
ACCGGCACAA GCAAACCCCA CCCACAACTT TTGCCCCATC AGAGACTTTT
TCTACTCAAC CAACTCAAGC ACCTGACATT AAGATTTCAA GTCAAGTGGA
GAGTTCTCTG GTTCCTACAG CTTGGGTGGA TAACACAGTT AATACCCCCA
AACAGTTGGA AATGGAGAAG AATGCAGAAC CCACATCCAA GGGAACACCA
CGGAGAAAAC ACGGGAAGAG GCCAAACAAA CATCGATATA CCCCTTCTAC
AGTGAGCTCA AGAGCGTCCG GATCCAAGCC CAGCCCTTCT CCAGAAAATA
AACATAGAAA CATTGTTACT CCCAGTTCAG AAACTATACT TTTGCCTAGA
ACTGTTTCTC TGAAAACTGA GGGCCCTTAT GATTCCTTAG ATTACATGAC
AACCACCAGA AAAATATATT CATCTTACCC TAAAGTCCAA GAGACACTTC
CAGTCACATA

TAAACCCACA TCAGATGGAA AAGAAATTAA GGATGATGTT GCCACAAATG
TTGACAAACA TAAAAGTGAC ATTTTAGTCA CTGGTGAATC AATTACTAAT
GCCATACCAA CTTCTCGCTC CTTGGTCTCC ACTATGGGAG AATTTAAGGA
AGAATCCTCT CCTGTAGGCT TTCCAGGAAC TCCAACCTGG AATCCCTCAA
GGACGGCCCA GCCTGGGAGG CTACAGACAG ACATACCTGT TACCACTTCT
GGGGAAAATC TTACAGACCC TCCCCTTCTT AAAGAGCTTG AGGATGTGGA
TTTCACTTCC GAGTTTTTGT CCTCTTTGAC AGTCTCCACA CCATTTCACC
AGGAAGAAGC TGGTTCTTCC ACAACTCTCT CAAGCATAAA AGTGGAGGTG
GCTTCAAGTC AGGCAGAAAC CACCACCCTT GATCAAGATC ATCTTGAAAC
CACTGTGGCT ATTCTCCTTT CTGAAACTAG ACCACAGAAT CACACCCCTA
CTGCTGCCCG GATGAAGGAG CCAGCATCCT CGTCCCCATC CACAATTCTC
ATGTCTTTGG GACAAACCAC CACCACTAAG CCAGCACTTC CCAGTCCAAG
AATATCTCAA GCATCTAGAG ATTCCAAGGA AAATGTTTTC TTGAATTATG
TGGGGAATCC AGAAACAGAA GCAACCCCAG TCAACAATGA AGGAACACAG
CATATGTCAG GGCCAAATGA ATTATCAACA CCCTCTTCCG ACCGGGATGC
ATTTAACTTG TCTACAAAGC TGGAATTGGA AAAGCAAGTA TTTGGTAGTA
GGAGTCTACC ACGTGGCCCA GATAGCCAAC GCCAGGATGG AAGAGTTCAT
GCTTCTCATC AACTAACCAG AGTCCCTGCC AAACCCATCC TACCAACAGC
AACAGTGAGG CTACCTGAAA TGTCCACACA AAGCGCTTCC AGATACTTTG
TAACTTCCCA GTCACCTCGT CACTGGACCA ACAAACCGGA AATAACTACA
TATCCTTCTG GGGCTTTGCC AGAGAACAAA CAGTTTACAA CTCCAAGATT
ATCAAGTACA ACAATTCCTC TCCCATTGCA CATGTCCAAA CCCAGCATTC
CTAGTAAGTT TACTGACCGA GAACTGACC AATTCAATGG TTACTCCAAA
GTGTTTGGAA ATAACAACAT CCCTGAGGCA AGAAACCCAG TTGGAAAGCC
TCCCAGTCCA AGAATTCCTC ATTATTCCAA TGGAAGACTC CCTTTCTTTA
CCAACAAGAC TCTTTCTTTT CCACAGTTGG GAGTCACCCG GAGACCCCAG
ATACCCACTT CTCCTGCCCC AGTAATGAGA GAGAGAAAAG TTATTCCAGG
TTCCTACAAC AGGATACATT CCCATAGCAC CTTCCATCTG GACTTTGGCC
CTCCGGCACC TCCGTTGTTG CACACTCCGC AGACCACGGG ATCACCCTCA
ACTAACTTAC AGAATATCCC TATGGTCTCT TCCACCCAGA GTTCTATCTC
CTTTATAACA TCTTCTGTCC AGTCCTCAGG AAGCTTCCAC CAGAGCAGCT
CAAAGTTCTT TGCAGGAGGA CCTCCTGCAT CCAAATTCTG GTCTCTTGGG
GAAAAGCCCC AAATCCTCAC CAAGTCCCCA CAGACTGTGT CCGTCACCGC
TGAGACAGAC ACTGTGTTCC CCTGTGAGGC AACAGGAAAA CCAAAGCCTT
TCGTTACTTG GACAAAGGTT TCCACAGGAG CTCTTATGAC TCCGAATACC
AGGATACAAC GGTTTGAGGT TCTCAAGAAC GGTACCTTAG TGATACGGAA
GGTTCAAGTA CAAGATCGAG GCCAGTATAT GTGCACCGCC AGCAACCTGC
ACGGCCTGGA CAGGATGGTG GTCTTGCTTT CGGTCACCGT GCAGCAACCT
CAAATCCTAG CCTCCCACTA CCAGGACGTC ACTGTCTACC TGGGAGACAC
CATTGCAATG GAGTGTCTGG CCAAAGGGAC CCCAGCCCCC CAAATTTCCT
GGATCTTCCC TGACAGGAGG GTGTGGCAAA CTGTGTCCCC

CGTGGAGAGC CGCATCACCC TGCACGAAAA CCGGACCCTT TCCATCAAGG
AGGCGTCCTT CTCAGACAGA GGCGTCTATA AGTGCGTGGC CAGCAATGCA
GCCGGGGCGG ACAGCCTGGC CATCCGCCTG CACGTGGCGG CACTGCCCCC
CGTTATCCAC CAGGAGAAGC TGGAGAACAT CTCGCTGCCC CCGGGGCTCA
GCATTCACAT TCACTGCACT GCCAAGGCTG CGCCCCTGCC CAGCGTGCGC
TGGGTGCTCG GGACGGTAC CCAGATCCGC CCCTCGCAGT TCCTCCACGG
GAACTTGTTT GTTTCCCCA ACGGGACGCT CTACATCCGC AACCTCGCGC
CCAAGGACAG CGGGCGCTAT GAGTGCGTGG CCGCCAACCT GGTAGGCTCC
GCGCGCAGGA CGGTGCAGCT GAACGTGCAG CGTGCAGCAG CCAACGCGCG
CATCACGGGC ACCTCCCCGC GGAGGACGGA CGTCAGGTAC GGAGGAACCC
TCAAGCTGGA CTGCAGCGCC TCGGGGGACC CCTGGCCGCG CATCCTCTGG
AGGCTGCCGT CCAAGAGGAT GATCGACGCG CTCTTCAGTT TTGATAGCAG
AATCAAGGTG TTTGCCAATG GGACCCTGGT GGTGAAATCA GTGACGGACA
AAGATGCCGG AGATTACCTG TGCGTAGCTC GAAATAAGGT TGGTGATGAC
TACGTGGTGC TCAAAGTGGA TGTGGTGATG AAACCGGCCA AGATTGAACA
CAAGGAGGAG AACGACCACA AAGTCTTCTA CGGGGGTGAC CTGAAAGTGG
ACTGTGTGGC CACCGGGCTT CCCAATCCCG AGATCTCCTG GAGCCTCCCA
GACGGGAGTC TGGTGAACTC CTTCATGCAG TCGGATGACA GCGGTGGACG
CACCAAGCGC TATGTCGTCT TCAACAATGG GACACTCTAC TTTAACGAAG
TGGGGATGAG GGAGGAAGGA GACTACACCT GCTTTGCTGA AAATCAGGTC
GGGAAGGACG AGATGAGAGT CAGAGTCAAG GTGGTGACAG CGCCCGCCAC
CATCCGGAAC AAGACTTACT GGCGGTTCA GGTGCCCTAT GGAGACGTGG
TCACTGTAGC CTGTGAGGCC AAAGGAGAAC CCATGCCCAA GGTGACTTGG
TTGTCCCCAA CCAACAAGGT GATCCCCACC TCCTCTGAGA AGTATCAGAT
ATACCAAGAT GGCACTCTCC TTATTCAGAA AGCCCAGCGT TCTGACAGCG
GCAACTACAC CTGCCTGGTC AGGAACAGCG CGGGAGAGGA TAGGAAGACG
GTGTGGATTC ACGTCAACGT CCAGCCACCC AAGATCAACG GTAACCCCAA
CCCCATCACC ACCGTGCGGG AGATAGCAGC CGGGGGCAGT CGGAAACTGA
TTGACTGCAA AGCTGAAGGC ATCCCCACCC CGAGGGTGTT ATGGGCTTTT
CCCGAGGGTG TGGTTCTGCC AGCTCCATAC TATGGAAACC GGATCACTGT
CCATGGCAAC GGTTCCCTGG ACATCAGGAG TTTGAGGAAG AGCGACTCCG
TCCAGCTGGT ATGCATGGCA CGCAACGAGG GAGGGGAGGC GAGGTTGATC
GTGCAGCTCA CTGTCCTGGA GCCCATGGAG AAACCCATCT TCCACGACCC
GATCAGCGAG AAGATCACGG CCATGGCGGG CCACACCATC AGCCTCAACT
GCTCTGCCGC GGGGACCCCG ACACCCAGCC TGGTGTGGGT CCTTCCCAAT
GGCACCGATC TGCAGAGTGG ACAGCAGCTG CAGCGCTTCT ACCACAAGGC
TGACGGCATG CTACACATTA GCGGTCTCTC CTCGGTGGAC GCTGGGGCCT
ACCGCTGCGT GGCCCGCAAT GCCGCTGGCC ACACGGAGAG GCTGGTCTCC
CTGAAGGTGG GACTGAAGCC AGAAGCAAAC AAGCAGTATC ATAACCTGGT
CAGCATCATC AATGGTGAGA CCCTGAAGCT CCCCTGCACC CCTCCCGGGG
CTGGGCAGGG ACGTTTCTCC TGGACGCTCC CCAATGGCAT GCATCTGGAG

GGCCCCCAAA CCCTGGGACG CGTTTCTCTT CTGGACAATG GCACCCTCAC
GGTTCGTGAG GCCTCGGTGT TTGACAGGGG TACCTATGTA TGCAGGATGG
AGACGGAGTA CGGCCCTTCG GTCACCAGCA TCCCCGTGAT TGTGATCGCC
TATCCTCCCC GGATCACCAG CGAGCCCACC CCGGTCATCT ACACCCGGCC
CGGGAACACC GTGAAACTGA ACTGCATGGC TATGGGGATT CCCAAAGCTG
ACATCACGTG GGAGTTACCG GATAAGTCGC ATCTGAAGGC AGGGGTTCAG
GCTCGTCTGT ATGGAAACAG ATTTCTTCAC CCCCAGGGAT CACTGACCAT
CCAGCATGCC ACACAGAGAG ATGCCGGCTT CTACAAGTGC ATGGCAAAAA
ACATTCTCGG CAGTGACTCC AAAACAACTT ACATCCACGT CTTCTGAAAT
GTGGATTCCA GAATGATTGC TTAGGAACTG ACAACAAAGC GGGGTTTGTA
AGGGAAGCCA GGTTGGGGAA TAGGAGCTCT TAAATAATGT GTCACAGTGC
ATGGTGGCCT CTGGTGGGTT TCAAGTTGAG GTTGATCTTG ATCTACAATT
GTTGGGAAAA GGAAGCAATG CAGACACGAG AAGGAGGGCT CAGCCTTGCT
GAGACACTTT CTTTTGTGTT TACATCATGC CAGGGGCTTC ATTCAGGGTG
TCTGTGCTCT GACTGCAATT TTTCTTCTTT TGCAAATGCC ACTCGACTGC
CTTCATAAGC GTCCATAGGA TATCTGAGGA ACATTCATCA AAAATAAGCC
ATAGACATGA ACAACACCTC ACTACCCCAT TGAAGACGCA TCACCTAGTT
AACCTGCTGC AGTTTTTACA TGATAGACTT TGTTCCAGAT TGACAAGTCA
TCTTTCAGTT ATTTCCTCTG TCACTTCAAA ACTCCAGCTT GCCCAATAAG
GATTTAGAAC CAGAGTGACT GATATATATA TATATATTTT AATTCAGAGT
TACATACATA CAGCTACCAT TTTATATGAA AAAGAAAAA CATTTCTTCC
TGGAACTCAC TTTTTATATA ATGTTTTATA TATATATTTT TTCCTTTCAA
ATCAGACGAT GAGACTAGAA GGAGAAATAC TTTCTGTCTT ATTAAAATTA
ATAAATTATT GGTCTTTACA AGACTTGGAT ACATTACAGC AGACATGGAA
ATATAATTTT AAAAAATTTC TCTCCAACCT CCTTCAAATT CAGTCACCAC
TGTTATATTA CCTTCTCCAG GAACCCTCCA GTGGGGAAGG CTGCGATATT
AGATTTCCTT GTATGCAAAG TTTTTGTTGA AAGCTGTGCT CAGAGGAGGT
GAGAGGAGAG GAAGGAGAAA ACTGCATCAT AACTTTACAG AATTGAATCT
AGAGTCTTCC CCGAAAAGCC CAGAAACTTC TCTGCAGTAT CTGGCTTGTC
CATCTGGTCT AAGGTGGCTG CTTCTTCCCC AGCCATGAGT CAGTTTGTGC
CCATGAATAA TACACGACCT GTTATTTCCA TGACTGCTTT ACTGTATTTT
TAAGGTCAAT ATACTGTACA TTTGATAATA AAATAATATT CTCCCAAAAA AAAAA

<u>Figure 53</u>   **Human OCP: nucleotide sequence of ORF**

ATGAAGGTAAAAGGCAGAGGAATCACCTGCTTGCTGGTCTCCTTTGCTGTGATCTGC
CTGGTCGCCACCCCTGGGGGCAAGGCCTGTCCTCGCCGCTGTGCCTGTTATATGCCT
ACGGAGGTACACTGCACATTTCGGTACCTGACTTCCATCCCAGACAGCATCCCGCCC
AATGTGGAACGCATCAATTTAGGATACAACAGCTTGGTTAGATTGATGGAAACAGAT
TTTTCTGGCCTGACCAAACTGGAGTTACTCATGCTTCACAGCAATGGCATTCACACA
ATCCCTGACAAGACCTTCTCAGATTTGCAGGCCTTGCAGGTCTTAAAAATGAGCTAT
AATAAAGTCCGAAAACTTCAGAAAGATACTTTTTATGGCCTCAGGAGCTTGACACGA
TTGCACATGGACCACAACAATATTGAGTTTATAAACCCAGAGGTTTTTTATGGGCTC
AACTTTCTCCGCCTGGTGCACTTGGAAGGAAATCAGCTCACTAAGCTCCACCCAGAT
ACATTTGTCTCTTTGAGCTACCTCCAGATATTTAAAATCTCTTTCATTAAGTTCCTA
TACTTGTCTGATAACTTCCTGACCTCCCTCCCTCAAGAGATGGTCTCCTATATGCCT
GACCTAGACAGCCTTTACCTGCATGGAAACCCATGGACCTGTGATTGCCATTTAAAG
TGGTTGTCTGACTGGATACAGGAGAAGCCAGATGTAATAAAATGCAAAAAAGATAGA
AGTCCCTCTAGTGCTCAGCAGTGTCCACTTTGCATGAACCCTAGGACTTCTAAAGGC
AAGCCGTTAGCTATGGTCTCAGCTGCAGCTTTCCAGTGTGCCAAGCCAACCATTGAC
TCATCCCTGAAATCAAAGAGCCTGACTATTCTGGAAGACAGTAGTTCTGCTTTCATC
TCTCCCCAAGGTTTCATGGCACCCTTTGGCTCCTCACTTTGAATATGACAGATCAG
TCTGGAAATGAAGCTAACATGGTCTGCAGTATTCAAAAGCCCTCAAGGACATCACCC
ATTGCATTCACTGAAGAAAATGACTACATCGTGCTAAATACTTCATTTTCAACATTT
TTGGTGTGCAACATAGATTACGGTCACATTCAGCCAGTGTGGCAAATTTTGGCTTTG
TACAGTGATTCTCCTCTGATACTAGAAAGGAGCCACTTGCTTAGTGAAACACCGCAG
CTCTATTACAAATATAAACAGGTGGCTCCTAAGCCTGAAGACATTTTTACCAACATA
GAGGCAGATCTCAGAGCAGATCCCTCTTGGTTAATGCAAGACCAAATTTCCTTGCAG
CTGAACAGAACTGCCACCACATTCAGTACATTACAGATCCAGTACTCCAGTGATGCT
CAAATCACTTTACCAAGAGCAGAGATGAGGCCAGTGAAACACAAATGGACTATGATT
TCAAGGGATAACAATACTAAGCTGGAACATACTGTCTTGGTAGGTGGAACCGTTGGC
CTGAACTGCCCAGGCCAAGGAGACCCCACCCCACACGTGGATTGGCTTCTAGCTGAT
GGAAGTAAAGTGAGAGCCCCTTATGTCAGTGAGGATGGACGGATCCTAATAGACAAA
AGTGGAAAATTGGAACTCCAGATGGCTGATAGTTTTGACACAGGCGTATATCACTGT
ATAAGCAGCAATTATGATGATGCAGATATTCTCACCTATAGGATAACTGTGGTAGAA
CCTTTGGTCGAAGCCTATCAGGAAAATGGGATTCATCACACAGTTTTCATTGGTGAA
ACACTTGATCTTCCATGCCATTCTACTGGTATCCCAGATGCCTCTATTAGCTGGGTT
ATTCCAGGAAACAATGTGCTCTATCAGTCATCAAGAGACAAGAAAGTTCTAAACAAT
GGCACATTAAGAATATTACAGGTCACCCCGAAAGACCAAGGTTATTATCGCTGTGTG
GCAGCCAACCCATCAGGGGTTGATTTTTTGATTTTCCAAGTTTCAGTCAAGATGAAA
GGACAAAGGCCCTTGGAGCATGATGGAGAAACAGAGGGATCTGGACTTGATGAGTCC
AATCCTATTGCTCATCTTAAGGAGCCACCAGGTGCACAACTCCGTACATCTGCTCTG
ATGGAGGCTGAGGTTGGAAAACACACCTCAAGCACAAGTAAGAGGCACAACTATCGG
GAATTAACACTCCAGCGACGTGGAGATTCAACACATCGACGTTTTAGGGAGAATAGG
AGGCATTTCCCTCCCTCTGCTAGGAGAATTGACCCACAACATTGGGCGGCACTGTTG
GAGAAAGCTAAAAAGAATGCTATGCCAGACAAGCGAGAAAATACCACAGTGAGCCCA
CCCCCAGTGGTCACCCAACTCCCAAACATACCTGGTGAAGAAGACGATTCCTCAGGC
ATGCTCGCTCTACATGAGGAATTTATGGTCCCGGCCACTAAAGCTTTGAACCTTCCA
GCAAGGACAGTGACTGCTGACTCCAGAACAATATCTGATAGTCCTATGACAAACATA
AATTATGGCACAGAATTCTCTCCTGTTGTGAATTCACAAATACTACCACCTGAAGAA
CCCACAGATTTCAAACTGTCTACTGCTATTAAAACTACAGCCATGTCAAAGAATATA

```
AACCCAACCATGTCAAGCCAAATACAAGGCACAACCAATCAACATTCATCCACTGTC
TTTCCACTGCTACTTGGAGCAACTGAATTTCAGGACTCTGACCAGATGGGAAGAGGA
AGAGAGCATTTCCAAAGTAGACCCCCAATAACAGTAAGGACTATGATCAAAGATGTC
AATGTCAAAATGCTTAGTAGCACCACCAACAAACTATTATTAGAGTCAGTAAATACC
ACAAATAGTCATCAGACATCTGTAAGAGAAGTGAGTGAACCCAGGCACAATCACTTC
TATTCTCACACTACTCAAATACTTAGCACCTCCACGTTCCCTTCAGATCCACACACA
GCTGCTCATTCTCAGTTTCCGATCCCTAGAAATAGTACAGTTAACATCCCGCTGTTC
AGACGCTTTGGGAGGCAGAGGAAAATTGGCGGAAGGGGGCGGATTATCAGCCCATAT
AGAACTCCAGTTCTGCGACGGCATAGATACAGCATTTTCAGGTCAACAACCAGAGGT
TCTTCTGAAAAAAGCACTACTGCATTCTCAGCCACAGTGCTCAATGTGACATGTCTG
TCCTGTCTTCCCAGGGAGAGGCTCACCACTGCCACAGCAGCATTGTCTTTTCCAAGT
GCTGCTCCCATCACCTTCCCCAAAGCTGACATTGCTAGAGTCCCATCAGAAGAGTCT
ACAACTCTAGTCCAGAATCCACTATTACTACTTGAGAACAAACCCAGTGTAGAGAAA
ACAACACCCACAATAAAATATTTCAGGACTGAAATTTCCCAAGTGACTCCAACTGGT
GCAGTCATGACATATGCTCCAACATCCATACCATGGAAAAAACTCACAAAGTAAAC
GCCAGTTACCCACGTGTGTCTAGCACCAATGAAGCTAAAAGAGATTCAGTGATTACA
TCGTCACTTTCAGGTGCTATCACCAAGCCACCAATGACTATTATAGCCATTACAAGG
TTTTCAAGAAGGAAAATTCCCTGGCAACAGAACTTTGTAAATAACCATAACCCAAAA
GGCAGATTAAGGAATCAACATAAAGTTAGTTTACAAAAAAGCACAGCTGTGATGCTT
CCTAAAACATCTCCTGCTTTACCACAGAGACAAAGTTCCCCTTTCCATTTCACCACA
CTTTCAACAAGTGTGATGCAAATTCCATCTAATACCTTGACTACCGCTCACCACACT
ACGACCAAAACACACAATCCTGGAAGTCTTCCAACAAAGAAGGAGCTTCCCTTCCCA
CCCCTTAACCCTATGCTTCCTAGTATTATAAGCAAAGACTCAAGTACAAAAAGCATC
ATATCAACGCAAACAGCAATACCAGCAACAACTCCTACCTTCCCTGCATCTGTCATC
ACTTATGAAACCCAAACAGAGAGATCTAGAGCACAAACAATACAAAGAGAACAGGAG
CCTCAAAAGAAGAACAGGACTGACCCAAACATCTCTCCAGACCAGAGTTCTGGCTTC
ACTACACCCACTGCTATGACACCTCCTGCTCTGGCATTCACTCATTCCCCACCAGAA
AACACAACTGGGATTTCAAGCACAATCAGTTTTCATTCAAGAACTCTTAATCTGACA
GATGTGATTGAAGAACTAGCCCAAGCAAGTACTCAGACTTTGAAGAGCACAATTGCT
TCTGAAACAACTTTGTCCAGCAAATCACACCAGAGTACCACAACTAGGAAAGCATCA
TTAGACACTCCCATACCACCATTCTTGAGCAGCAGTGCTACTCTAATGCCAGTTCCC
ATCTCCCCTCCCTTTACTCAGAGAGCAGTTACTGACACACGTGGCGACTCCCATTTC
CGGCTTATGACAAATACAGTGGTCAAGCTGCACGAATCCTCAAGGCACAATCTCCAA
ATGCCAAGTTCACAATTGGAACCACTCACTTCATCTACCTCTAATCTGTTACATTCT
ACTCCCATGCCAGCACTAACAACAGTTAAATCACAGAATTCCAAATTAACTCCATCT
CCCTGGGCAGAATACCAATTTTGGCACAAACCATACTCAGACATTGCTGAAAAAGGC
AAAAAGCCAGAAGTAAGCATGTTGGCTACTACAGGCCTGTCCGAGGCCACCACTCTT
GTTTCAGATTGGGATGGACAGAAGAACACAAAGAAGAGTGACTTTGATAAGAAACCA
GTTCAAGAAGCAACAACTTCCAAACTCCTTCCCTTTGACTCTTTGTCTAGGTATATA
TTTGAAAAGCCCAGGATAGTTGGAGGAAAAGCTGCAAGTTTTACTATTCCAGCTAAC
TCAGATGCCTTTCTTCCCTGTGAAGCTGTTGGAAATCCCCTGCCCACCATTCATTGG
ACCAGAGTTTCAGGACTTGATTTATCTAGAGGAAACCAGAATAGCAGGGTCCAGGTT
CTCCCCAATGGTACCCTGTCCATCCAGAGGGTGGAAATTCAGGACCGCGGACAGTAC
TTGTGTTCCGCATCCAATCTGTTTGGCACAGACCACCTTCATGTCACCTTGTCTGTG
GTTTCCTATCCTCCCAGGATCCTGGAGAGACGTACCAAAGAGATCACAGTTCATTCC
GGAAGCACTGTGGAACTGAAGTGCAGAGCAGAAGGTAGGCCAAGCCCTACAGTTACC
TGGATTCTTGCAAACCAAACAGTTGTCTCAGAATCATCCCAGGGAAGTAGGCAGGCT
GTGGTGACGGTTGACGGAACATTGGTCCTCCACAATCTCAGTATTTATGACCGTGGC
TTTTACAAATGTGTGGCCAGCAACCCAGGTGGCCAGGATTCACTGCTGGTTAAAATA
CAAGTCATTGCAGCACCACCTGTTATTCTAGAGCAAAGGAGGCAAGTCATTGTAGGC
ACTTGGGGTGAAAGTTTAAAACTGCCCTGTACTGCAAAAGGAACTCCTCAGCCCAGC
```

GTTTACTGGGTCCTCTCTGATGGCACTGAAGTGAAACCATTACAGTTTACCAATTCC
AAGTTGTTCTTATTTTCAAATGGGACTTTGTATATAAGAAACCTAGCCTCTTCAGAC
AGGGGCACTTATGAATGCATTGCTACCAGTTCCACTGGTTCGGAGCGAAGAGTAGTA
ATGCTTACAATGGAAGAGCGAGTGACCAGCCCCAGGATAGAAGCTGCATCCCAGAAA
AGGACTGAAGTGAATTTTGGGGACAAATTACTACTGAACTGCTCAGCCACTGGGGAG
CCCAAACCCCAAATAATGTGGAGGTTACCATCCAAGGCTGTGGTCGACCAGTGGAGC
TGGATCCACGTCTACCCTAATGGATCCCTGTTTATTGGATCAGTAACAGAAAAAGAC
AGTGGTGTCTACTTGTGTGTGGCAAGAAACAAAATGGGGGATGATCTGATACTGATG
CATGTTAGCCTAAGACTGAAACCTGCCAAAATTGACCACAAGCAGTATTTTAGAAAG
CAAGTGCTCCATGGGAAAGATTTCCAAGTAGATTGCAAAGCTTCCGGCTCCCCAGTG
CCAGAGATATCTTGGAGTTTGCCTGATGGAACCATGATCAACAATGCAATGCAAGCC
GATGACAGTGGCCACAGGACTAGGAGATATACCCTTTTCAACAATGGAACTTTATAC
TTCAACAAAGTTGGGGTAGCGGAGGAAGGAGATTATACTTGCTATGCCCAGAACACC
CTAGGGAAAGATGAAATGAAGGTCCACTTAACAGTTATAACAGCTGCTCCCCGGATA
AGGCAGAGTAACAAAACCAACAAGAGAATCAAAGCTGGAGACACAGCTGTCCTTGAC
TGTGAGGTCACTGGGGATCCCAAACCAAAAATATTTTGGTTGCTGCCTTCCAATGAC
ATGATTTCCTTCTCCATTGATAGGTACACATTTCATGCCAATGGGTCTTTGACCATC
AACAAAGTGAAACTGCTCGATTCTGGAGAGTACGTATGTGTAGCCCGAAATCCCAGT
GGGGATGACACCAAAATGTACAAACTGGATGTGGTCTCTAAACCTCCATTAATCAAT
GGTCTGTATACAAACAGAACTGTTATTAAAGCCACAGCTGTGAGACATTCCAAAAAA
CACTTTGACTGCAGAGCTGAAGGGACACCATCTCCTGAAGTCATGTGGATCATGCCA
GACAATATTTTCCTCACAGCCCCATACTATGGAAGCAGAATCACAGTCCATAAAAAT
GGAACCTTGGAAATTAGGAATGTGAGGCTTTCAGATTCAGCCGACTTTATCTGTGTG
GCCCGAAATGAAGGTGGAGAGAGCGTGTTGGTAGTACAGTTAGAAGTACTGGAAATG
CTGAGAAGACCGACATTTAGAAATCCATTTAATGAAAAAATAGTTGCCCAGCTGGGA
AAGTCCACAGCATTGAATTGCTCTGTTGATGGTAACCCACCACCTGAAATAATCTGG
ATTTTACCAAATGGCACACGATTTTCCAATGGACCACAAAGTTATCAGTATCTGATA
GCAAGCAATGGTTCTTTTATCATTTCTAAAACAACTCGGGAGGATGCAGGAAAATAT
CGCTGTGCAGCTAGGAATAAAGTTGGCTATATTGAGAAATTAGTCATATTAGAAATT
GGCCAGAAGCCAGTTATTCTTACCTATGCACCAGGGACAGTAAAAGGCATCAGTGGA
GAATCTCTATCACTGCATTGTGTGTCTGATGGAATCCCTAAGCCAAATATCAAATGG
ACTATGCCAAGTGGTTATGTAGTAGACAGGCCTCAAATTAATGGGAAATACATATTG
CATGACAATGGCACCTTAGTCATTAAAGAAGCAACAGCTTATGACAGAGGAAACTAT
ATCTGTAAGGCTCAAAATAGTGTTGGTCATACACTGATTACTGTTCCAGTAATGATT
GTAGCCTACCCTCCCCGAATTACAAATCGTCCACCCAGGAGTATTGTCACCAGGACA
GGGGCAGCCTTTCAGCTCCACTGTGTGGCCTTGGGAGTTCCCAAGCCAGAAATCACA
TGGAGATGCCTGACCACTCCCTTCTCTCAACGGCAAGTAAAGAGAGGACACATGGA
AGTGAGCAGCTTCACTTACAAGGTACCCTAGTCATTCAGAATCCCCAAACCTCCGAT
TCTGGGATATACAAATGCACAGCAAGAACCCACTTGGTAGTGATTATGCAGCAACG
TATATTCAAGTAATCTGA

<u>Figure 54</u>   **Human OCP: predicted amino acid sequence**

```
MKVKGRGITC LLVSFAVICL VATPGGKACP RRCACYMPTE VHCTFRYLTS
IPDSIPPNVE RINLGYNSLV RLMETDFSGL TKLELLMLHS NGIHTIPDKT
FSDLQALQVL KMSYNKVRKL QKDTFYGLRS LTRLHMDHNN IEFINPEVFY
GLNFLRLVHL EGNQLTKLHP DTFVSLSYLQ IFKISFIKFL YLSDNFLTSL
PQEMVSYMPD LDSLYLHGNP WTCDCHLKWL SDWIQEKPDV IKCKKDRSPS
SAQQCPLCMN PRTSKGKPLA MVSAAAFQCA KPTIDSSLKS KSLTILEDSS
SAFISPQGFM APFGSLTLNM TDQSGNEANM VCSIQKPSRT SPIAFTEEND
YIVLNTSFST FLVCNIDYGH IQPVWQILAL YSDSPLILER SHLLSETPQL
YYKYKQVAPK PEDIFTNIEA DLRADPSWLM QDQISLQLNR TATTFSTLQI
QYSSDAQITL PRAEMRPVKH KWTMISRDNN TKLEHTVLVG GTVGLNCPGQ
GDPTPHVDWL LADGSKVRAP YVSEDGRILI DKSGKLELQM ADSFDTGVYH
CISSNYDDAD ILTYRITVVE PLVEAYQENG IHHTVFIGET LDLPCHSTGI
PDASISWVIP GNNVLYQSSR DKKVLNNGTL RILQVTPKDQ GYYRCVAANP
SGVDFLIFQV SVKMKGQRPL EHDGETEGSG LDESNPIAHL KEPPGAQLRT
SALMEAEVGK HTSSTSKRHN YRELTLQRRG DSTHRRFREN RRHFPPSARR
IDPQHWAALL EKAKKNAMPD KRENTTVSPP PVVTQLPNIP GEEDDSSGML
ALHEEFMVPA TKALNLPART VTADSRTISD SPMTNINYGT EFSPVVNSQI
LPPEEPTDFK LSTAIKTTAM SKNINPTMSS QIQGTTNQHS STVFPLLLGA
TEFQDSDQMG RGREHFQSRP PITVRTMIKD VNVKMLSSTT NKLLLESVNT
TNSHQTSVRE VSEPRHNHFY SHTTQILSTS TFPSDPHTAA HSQFPIPRNS
TVNIPLFRRF GRQRKIGGRG RIISPYRTPV LRRHRYSIFR STTRGSSEKS
TTAFSATVLN VTCLSCLPRE RLTTATAALS FPSAAPITFP KADIARVPSE
ESTTLVQNPL LLLENKPSVE KTTPTIKYFR TEISQVTPTG AVMTYAPTSI
PMEKTHKVNA SYPRVSSTNE AKRDSVITSS LSGAITKPPM TIIAITRFSR
RKIPWQQNFV NNHNPKGRLR NQHKVSLQKS TAVMLPKTSP ALPQRQSSPF
HFTTLSTSVM QIPSNTLTTA HHTTTKTHNP GSLPTKKELP FPPLNPMLPS
IISKDSSTKS IISTQTAIPA TTPTFPASVI TYETQTERSR AQTIQREQEP
QKKNRTDPNI SPDQSSGFTT PTAMTPPALA FTHSPPENTT GISSTISFHS
RTLNLTDVIE ELAQASTQTL KSTIASETTL SSKSHQSTTT RKASLDTPIP
PFLSSSATLM PVPISPPFTQ RAVTDTRGDS HFRLMTNTVV KLHESSRHNL
QMPSSQLEPL TSSTSNLLHS TPMPALTTVK SQNSKLTPSP WAEYQFWHKP
YSDIAEKGKK PEVSMLATTG LSEATTLVSD WDGQKNTKKS DFDKKPVQEA
TTSKLLPFDS LSRYIFEKPR IVGGKAASFT IPANSDAFLP CEAVGNPLPT
IHWTRVSGLD LSRGNQNSRV QVLPNGTLSI QRVEIQDRGQ YLCSASNLFG
TDHLHVTLSV VSYPPRILER RTKEITVHSG STVELKCRAE GRPSPTVTWI
LANQTVVSES SQGSRQAVVT VDGTLVLHNL SIYDRGFYKC VASNPGGQDS
LLVKIQVIAA PPVILEQRRQ VIVGTWGESL KLPCTAKGTP QPSVYWVLSD
GTEVKPLQFT NSKLFLFSNG TLYIRNLASS DRGTYECIAT SSTGSERRVV
MLTMEERVTS PRIEAASQKR TEVNFGDKLL LNCSATGEPK PQIMWRLPSK
AVVDQWSWIH VYPNGSLFIG SVTEKDSGVY LCVARNKMGD DLILMHVSLR
LKPAKIDHKQ YFRKQVLHGK DFQVDCKASG SPVPEISWSL PDGTMINNAM
QADDSGHRTR RYTLFNNGTL YFNKVGVAEE GDYTCYAQNT LGKDEMKVHL
TVITAAPRIR QSNKTNKRIK AGDTAVLDCE VTGDPKPKIF WLLPSNDMIS
FSIDRYTFHA NGSLTINKVK LLDSGEYVCV ARNPSGDDTK MYKLDVVSKP
PLINGLYTNR TVIKATAVRH SKKHFDCRAE GTPSPEVMWI MPDNIFLTAP
YYGSRITVHK NGTLEIRNVR LSDSADFICV ARNEGGESVL VVQLEVLEML
RRPTFRNPFN EKIVAQLGKS TALNCSVDGN PPPEIIWILP NGTRFSNGPQ
SYQYLIASNG SFIISKTTRE DAGKYRCAAR NKVGYIEKLV ILEIGQKPVI
```

```
LTYAPGTVKG ISGESLSLHC VSDGIPKPNI KWTMPSGYVV DRPQINGKYI
LHDNGTLVIK EATAYDRGNY ICKAQNSVGH TLITVPVMIV AYPPRITNRP
PRSIVTRTGA AFQLHCVALG VPKPEITWEM PDHSLLSTAS KERTHGSEQL
HLQGTLVIQN PQTSDSGIYK CTAKNPLGSD YAATYIQVI*
```